

Measuring Comprehension of Abstract Data Visualisations

Mark Shovman

A thesis for the degree of Doctor of Philosophy at



2011

This item is protected by original copyright.

Mark M. Shovman

MEASURING THE COMPREHENSION OF
ABSTRACT DATA VISUALISATIONS

A thesis submitted in partial fulfilment of the
requirements for the degree of

PhD

University of Abertay Dundee

Approved by _____

Chairman of Supervisory Committee

Programme Authorised to Offer Degree _____

Date _____

Abstract

Common visualisation techniques such as bar-charts and scatter-plots are not sufficient for visual analysis of large sets of complex multidimensional data. Technological advancements have led to a proliferation of novel visualisation tools and techniques that attempt to meet this need. A crucial requirement for efficient visualisation tool design is the development of objective criteria for visualisation quality, informed by research in human perception and cognition.

This thesis presents a multidisciplinary approach to address this requirement, underpinning the design and implementation of visualisation software with the theory and methodology of cognitive science. An opening survey of visualisation practices in the research environment identifies three primary uses of visualisations: the detection of outliers, the detection of clusters and the detection of trends. This finding, in turn, leads to a formulation of a cognitive account of the visualisation comprehension processes, founded upon established theories of visual perception and reading comprehension. Finally, a psychophysical methodology for objectively assessing visualisation efficiency is developed and used to test the efficiency of a specific visualisation technique, namely an interactive three-dimensional scatterplot, in a series of four experiments.

The outcomes of the empirical study are three-fold. On a concrete applicable level, three-dimensional scatterplots are found to be efficient in trend detection but not in outlier detection. On a methodological level, ‘pop-out’ methodology is shown to be suitable for assessing visualisation efficiency. On a theoretical level, the cognitive account of visualisation comprehension processes is enhanced by empirical findings, e.g. the significance of the learning curve parameters. All these provide a contribution to a ‘science of visualisation’ as a coherent scientific paradigm, both benefiting fundamental science and meeting an applied need.

Table of Contents

Abstract	i
Table of Contents	ii
List of Figures	iii
List of Tables.....	vii
Enclosed CD Contents.....	vii
Acknowledgements	viii
Chapter 1. Introduction.....	1
Chapter 2. Visualisation research: a patchwork paradigm	6
Chapter 3. Preliminary study: visualisation practices in research	17
Chapter 4. Towards a theory of visualisation comprehension.....	23
Chapter 5. Methodology for visualisation evaluation	32
Chapter 6. Experiment 1: outlier detection in a 3D scatterplot	48
Chapter 7. Experiment 2: the role of interactive view control in outlier detection	58
Chapter 8. Experiment 3: roles of high and low-level objects in trend detection.....	85
Chapter 9. Experiment 4: 3D view control in trend detection.....	108
Chapter 10. Discussion and conclusions	131
Appendix: abstracts of peer-reviewed publications.....	i
References	iv

List of Figures

Figure 1 – Anscombe’s quartet (Anscombe, 1973) is an example of how plotting the data can improve the understanding of underlying regularities. The four datasets presented have identical statistical properties, but widely differing patterns.....	2
Figure 2 – a model of processes involved in visualisation comprehension. Elements that were added by the author and are not commonly present in literature are marked by dashed lines and italics.	14
Figure 3 – a schematic stimulus from the tree-tracing task. In a typical task, two intertwined 3D structures are presented in 2D, stereo, or head-tracked movement, and participants are asked to identify the structure to which the highlighted leaf belongs.	15
Figure 4 – outliers, clusters and trends in 1D and 2D visualisation. Top: 1D histogram; bottom: 2D scatterplot. Left: the visualisation; right: schematic representation of parsing into outliers, clusters and trends. Outliers are denoted by X, clusters by solid lines and trends by dashed lines.	21
Figure 5 – optical illusion combining effects of grouping, figure-ground segregation, and illusory contours to create a perception of three-dimensional white cube from sensation of 24 black sector shapes.	24
Figure 6 – main application object design. Boxes: classes; rounded boxes: Java library classes; wavy boxes: files.....	44
Figure 7 – Experiment #3 MainForm dialog, showing fields for gathering general data about the participant, as well as the controls to run specific tests.	45
Figure 8 – file structure for experiment resources and logs. Gray boxes: folders; white boxes: files.....	47
Figure 9 – stimuli used in Experiment #1. Top row: dataset with 25 points; bottom row: dataset with 55 points. Columns A, B & C show different stages of rotation: -22.5°, 0° and 22.5° respectively. Column D presents a side view, rotated 90° around vertical axis, which clearly shows the outlying target. This side view was never shown to the participants.....	52
Figure 10 – results of Experiment #1. A: average response times for correct answers; B: error rate. Error bars indicate ± 1 standard error of mean; solid lines and filled circles: target at 5 SD from the mean; dashed lines and empty circles: target at 3 SD from the mean. Greyed dashed line and pluses in chart B is the theoretical chance performance level. Lines are at linear fit to data. N=25.....	53
Figure 11 – estimated marginal means of response times for correct answers after correcting for target size and distance from centre. Error bars indicate ± 1 standard error of mean; solid lines and filled circles: target at 5 SD from the mean; dashed lines and empty circles: target at 3 SD from the mean. Lines are at linear fit to data. N=25.....	55
Figure 12 – schematic position of the observer’s view point (symbol of an eye), the stimulus reference plane (line) and the target (circle), in Experiment #1 (left); Experiment #2 ‘floor’ slant (middle) and ‘ceiling’ slant (right); side view.	59
Figure 13 – handedness questionnaire form used in Experiment #2.....	63
Figure 14 – Experiment #2 interface buttons.	64

Figure 15 – stimuli used in Experiment #2; a dataset with 343 points with a position outlier, in ‘floor’ slant, in a manual rotation mode, showing left and right rotation limits. 66

Figure 16 – stimuli used in Experiment #2. Left: side view of the dataset shown in previous figure, with the position outlier clearly visible; right: a dataset with 343 points and a colour outlier..... 67

Figure 17 – mouse movements during colour outlier trials; overlay of all trials. Blue dots: mouse movements without button press; red dots: button presses and mouse movements with button pressed (dragging). The axes are screen XY coordinates in pixels..... 69

Figure 18 – mouse movements during position outlier trials with passive view control mode; overlay of all trials. Top: screen coordinates view (the axes are screen XY coordinates in pixels); bottom: XxYxTime view (X and Y axes are screen coordinates in pixels; Z axis in time from the start of the trial in milliseconds) Blue dots: mouse movements without button press; red dots: button presses and mouse movements with button pressed (dragging)..... 71

Figure 19 – mouse movements during position outlier trials with active view control mode; overlay of all trials. Top: screen coordinates view (the axes are screen XY coordinates in pixels); bottom: XxYxTime view (X and Y axes are screen coordinates in pixels; Z axis in time from the start of the trial in milliseconds) Blue dots: mouse movements without button press; red dots: button presses and mouse movements with button pressed (dragging)..... 72

Figure 20 – Percent of correct answers. Chance level is the inverse of the number of points. 73

Figure 21 – P-P plots of interaction time. Top: passive rotation; bottom: active rotation; left: no transformation, right: natural logarithm link function. 73

Figure 22 – effect of data set size on total response time in trials with colour outlier 74

Figure 23 – effects of data set size on pre- and post-interaction time. Left: pre-interaction (wait) time; Right: post-interaction (response) time. Error bars indicate ± 1 standard error of mean. 74

Figure 24 – effects of data set size on interaction time. Error bars indicate ± 1 standard error of mean. 75

Figure 25 – A: effects of reference plane slant on interaction time. Left: ‘floor’ slant (bottom closer to the viewer); right: ‘ceiling’ slant (top closer to the viewer). B: effects of target colour on interaction time. Error bars indicate ± 1 standard error of mean..... 77

Figure 26 – effects of order of presentation on interaction time. X-axis: trial blocks in order of presentation. Dotted line: trend for active rotation (the trend for passive rotation was not significant). Error bars indicate ± 1 standard error of mean..... 78

Figure 27 – change in interaction time over the course of Experiment #2, for active interactivity trials only..... 80

Figure 28 – juxtaposition of personal changes in strategy and changes in performance. A: changes in sweep range; B: changes in sweep speed. Dotted line in A signifies linear trend. 81

Figure 29 - 3D structures used in Experiment #3..... 90

Figure 30 – sample Experiment #3 stimuli. Left panel: two datasets of 64 points. Middle panel: three datasets of 144 points. Right panel: four datasets of 256 points..... 91

Figure 31 – schematic outline of mouse log parsing algorithm. Bold dotted line: mouse movements; solid double-arrow lines: meaningful time segments.....	92
Figure 32 – histograms of net rotation time distribution. Left panel depicts data with no link function; right panel indicates data after applying a natural logarithm link function.....	93
Figure 33 – Frequency distribution of correct answers made without scene rotation, by participants. Participants who never responded without scene rotation are not shown.....	94
Figure 34 - per-participant performance: speed and accuracy. Chance rate for accuracy is 36%. Error bars indicate ± 1 standard error of mean.....	95
Figure 35 – differences in performance (speed and accuracy) for different target data set formulae. Error bars indicate ± 1 standard error of mean.....	96
Figure 36 – By participant bar charts of accuracy (top) and speed (bottom) of possible green-yellow confusion. Error bars in bottom chart indicate ± 1 standard error of mean. Filled symbols indicate possible green/yellow confusion, open symbols indicate no confusion.	97
Figure 37 – Bar chart depicting proportion of Green/Yellow confusion in overall wrong answers for each participant in descending order. Horizontal gridlines show chance level (17%).	98
Figure 38 – net interaction time (seconds) as a function of number of points per stimulus data set, for the three conditions (2 data sets, 3 data sets and 4 data sets respectively). Error bars indicate ± 1 standard error of mean.	99
Figure 39 – trial sequence charts depicting changes in performance during the course of experiment. Top panel depicts accuracy by trial order; bottom: speed. Vertical gridlines denote blocks. Dotted line depicts predicted slope. Error bars in the bottom chart indicate ± 1 standard error of mean.....	100
Figure 40 – Frequency distribution of average mouse sweep time (milliseconds). Left panel depicts untransformed data; right hand panel depicts data following a logarithm transform.	101
Figure 41 – Scatterplot depicting Digit Span as a function of participant age. Participant initials are adjacent to data points.	103
Figure 42 – Scatterplot showing correlation between digit span and the influence of number of datapoint groups on interaction time. The influence is calculated as the uncorrected slope coefficient from regression analysis.	104
Figure 43 – Scatterplots showing correlations of age and digit span with the effect of number of data points per group on interaction time. Left panel depicts relationship between interaction time and digit span, right panel depicts interaction time influence and age in years. Note: in the second graph, removing data of participant AI does not change the trend.	104
Figure 44 – an InterSense IS-900 wand, used in Experiment #4.....	109
Figure 45 – some of the 3D structures used as targets in Experiment #4.....	114
Figure 46 – distribution of response times in Experiment #4.....	117
Figure 47 – per-participant performance summary; standardized scores. Top left: speed and accuracy; top right: pen and paper tests; bottom: average of the four Z-scores. Vertical	

dotted line in the top left chart is at 36% which is the chance performance level for this task (the same as in Experiment #3). 118

Figure 48 – differences in speed and accuracy caused by target set structure (left) and colour (right). Error bars indicate ± 1 standard error of mean..... 119

Figure 49 – relative proportion of trials that were answered correctly in either 1DOF condition, 3DOF condition, or both. 121

Figure 50 – accuracy (left) and response time (right) by rotation type and block order. Error bars on the right chart indicate ± 1 standard error of mean..... 123

Figure 51 – response time for correct answers per participant per rotation mode. Left: participants starting from 1DOF condition; right: participants starting with 3DOF condition. Light bars: 1DOF rotation; dark bars: 3DOF rotation. Error bars indicate ± 1 standard error of mean. X axis ordered by the difference in response time between first and second block.124

Figure 52 – effect of number of point groups on response time, represented by b parameter in the regression analysis. Error bars indicate ± 1 standard error of mean. X-axis ordered by average value of 1DOF and 3DOF condition..... 125

Figure 53 – 2D histograms of scene rotation per participant. First and fourth row: 3DOF; second and third: 1DOF; first and third row: block 1; second and forth: block 2. 129

List of Tables

Table 1 – results of regression analysis of wait, interaction and response time as a function of data set size, separately for active and passive rotation modes. Only trials with position outlier, and only trials were participants gave the correct answer were analysed.	75
Table 2 – regression analysis of training effects	77
Table 3 – Experiment #2 enhanced main analysis, linear regression coefficients (b slope, β standardized slope and significance). ‘Floor’ slant was encoded as 1 and ‘ceiling’ slant as 2.78	
Table 4 – Frequency distribution of correct answers made without scene rotation, grouped by target data set formula coefficients.....	94
Table 5 – linear regression coefficients (b slope, β standardized slope and significance respectively) for Experiment #3 main analysis.	99
Table 6 – Pearson r of strategy indicators with interaction time.....	101
Table 7 – regression coefficients of Experiment #3 augmented main analysis.....	102
Table 8 – Experiment #4 main analysis: regression coefficients.	121
Table 9 – linear regression coefficients for the full main analysis of Experiment #4.	122
Table 10 – Experiment #4 main analysis coefficients, separately for block order and interactivity mode.....	125

Enclosed CD Contents

Digital copy of the thesis

Digital copies of peer-reviewed presentations

Scan of Ethical Approval form

Paper-based resources for each experiment such as questionnaires

Raw collected data in .csv format

Acknowledgements

בס"ד

My thanks go, first and foremost, to my family, near and far, for their patience and support; to my supervisors, Kenneth Scott-Brown, James Bown and Andrea Szymkowiak, for showing me the ropes; and to fellow students from the White Space Research group for being a good company. Colin Roscoe deserves a special mention for his merciless proofreading.

I would like to thank the members of Scottish Vision Group and the Fife-Tayside Vision Meetings: Madeleine Keehner, Vit Drga, Julie Harris, Ben Vincent, Michael Swanston and especially David Heeley, for insightful comments, rewarding discussions, support and encouragement.

Parts of this thesis were presented at the European Conference of Visual Perception 2008, Information Visualisation 2009, meetings of Applied Vision Association 2010, Psychology Postgraduate Action Group 2008, Scottish Vision Group 2008, 2009 and 2010, and the meetings of the Fife-Tayside Vision Group. A full list of peer-reviewed publications, including abstracts, can be seen in the appendices.

Chapter 1. Introduction

Historically, Visual Analytics is thought to begin with the works of William Playfair (Friendly & Denis, 2001). In his *Commercial and Political Atlas*, published in 1786, he introduces most of the graphical forms used to present data to this day: bar charts, pie charts and time-series plots. Surprisingly for modern readers, Playfair felt he has to spend some effort convincing his audience to use these. In a preface to his work, he writes as follows: *"The reader will find, five minutes attention to the principle on which they (the charts) are constructed, a saving of much labour and time; but, without that trifling attention, he may as well look at a blank sheet of paper as at one of the Charts"* (Playfair, 1801). In fact, the idea of presenting abstract numerical data in visual form can be traced back to Descartes' invention of analytical geometry in the first half of the XVII century. However, Descartes only applied this method to mathematical domain, using it, to use his own expression, to *'solve geometry with algebra'* (as cited in Friendly & Denis, 2001). It is not until Playfair's work that charts started to be used in an applied context.

In the most general terms, visualisation is the process of creating images, diagrams, or animations to convey information, as well as the result of such a process. In this thesis, for the sake of narrative flow, the terms *chart*, *graph*, *plot* and *visualisation* will be used interchangeably to refer to any kind of visual presentation of numerical data that is not in itself purely textual or numerical. Specific visualisation techniques can vary greatly depending on the target audience, the nature of the information, and the medium used. One example of a technique change to suit different target audience is the difference between a newspaper weather map that aims to inform the general public about the weather and a satellite image overlay that helps meteorologists to generate weather forecasts.

In abstract, mathematical terms a visualisation technique can be defined as a function that maps from data space to image space, for example, substituting colour for temperature, left-to-right for past-to-future and so on (Mackinlay, 1986). In many applications, the target space

is not purely visual, containing components of other, usually auditory or haptic, modalities. These multi-modal visualisations, or *perceptualisations* as they are sometimes called (Grinstein & Smith, 1990), are outside the scope of this work and are mentioned here only for the sake of completeness of the argument.

In more than two centuries that have passed since Playfair’s *Atlas*, plots, charts and graphs have become an indispensable part of our cultural expressive vocabulary. Charts are used in science, education, politics, mass-media, law and art. In education, charts help achieve understanding and long-term retention of the material (Shah & Hoeffner, 2002). In mass-media, ‘infographics’ (McCandless, 2009) are designed to be visually engaging and aesthetically appealing as well as informative. In law, the charts are carefully prepared to “...capture a juror’s attention, enlighten them and facilitate a decision in your favour.” (Arkfeld, 2001).

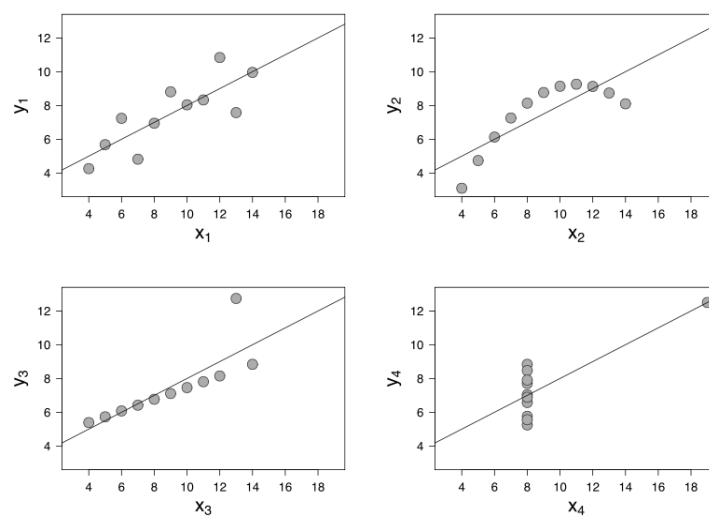


Figure 1 – Anscombe’s quartet (Anscombe, 1973) is an example of how plotting the data can improve the understanding of underlying regularities. The four datasets presented have identical statistical properties, but widely differing patterns.

In all these cases, however, the message that the chart is to communicate is known beforehand. The most challenging use case and also, arguably, the one most in need of further research, is when the message is unknown. Then, visualisations are used to generate insights and acquire novel information, extending the boundaries of human knowledge. In

statistics this use of visualisations, e.g. plotting the raw data as a histogram or scatterplot, is a part of exploratory data analysis (Tukey, 1977), and is considered an indispensable first step in working with data, preceding any formal statistical tests (see e.g. Figure 1). An iconic example of exploratory visualisation is the discovery of the 3D structure of DNA (Watson, 2001): only after soldering models of amino-acids together, were the researchers able to see the beautifully simple helical structure.

Since the computer revolution, the introduction of increasingly powerful graphical hardware and software has been opening new possibilities for data visualisations: fast, dynamic, multi-modal and immersive (Chaomei Chen, 2004). The demand for such tools is high, especially in high-stakes fields such as bio-medical research, security and finance (Thomas & Cook, 2005). Increasingly, in industry, academia, military etc., dedicated large-scale Virtual Reality facilities are set up for collaborative visual data analysis (e.g. Lidal, Langeland, & Giertsen, 2007). The Human Interactive Visual Environment laboratory (HIVE) in the University of Abertay is such a facility, presenting a unique opportunity to study visualisations generated with cutting-edge hardware and software.

Unfortunately, the ubiquity of visualisations inevitably leads to diversity in quality. In the printing industry, ‘best practices’ in chart design, stemming from writers’ and editors’ experience, have evolved over the years (Bigwood & Spore, 2003). The implementation of these practices, though, has not been overly widespread: the guides on data presentation are full of real-world examples that violate them; that are at best unreadable, and at worst misleading. Also, a significant proportion of these best practices are dictated by the restrictions of the printing medium: static, two-dimensional and non-interactive.

Computer-based visualisation is a novel and fast-developing field. However, the criteria behind creating the new visualisations often seem to be based on the subjective preferences of the authors (Thomas & Cook, 2005). Best practices take longer to evolve than it takes for novel visualisations to appear, and so the quality issues are more prominent. The strong

point of the visualisation, its visual power, is frequently counterproductive: inefficient visualisations can be visually quite engaging, and this makes spotting their inefficiency difficult. As a result, many novel tools are developed but never used, with users reverting to ‘tried and true’ static two-dimensional techniques such as bar- or pie-charts (Nesbitt, 2004).

One often-proposed solution to the problem of proliferation of substandard visualisation tools is a framework of systematic quality evaluation. Several reviews of approaches to such an evaluation will be presented and analysed in Chapter 2 of this thesis. The summary of the review is that evaluation must rest on a theoretical foundation, and, while there have been several attempts to propose a theory of graph comprehension (Cleveland & McGill, 1984; Kosslyn, 2006; Nesbitt, 2004; Pinker, 1990; Van Wijk, 2005; Ware, 2000) there is, at the moment, no comprehensive account of the processes taking place between the chart and the viewer: the processes of visualisation comprehension.

It has been noted that “*Graphs are a recent invention, and if they are an especially effective method of communication, it is because they exploit general cognitive and perceptual mechanisms effectively*” (Pinker, 1990, p. 73). A similar line of reasoning has been made regarding reading comprehension, a much better researched area, and has received ample supporting evidence (Kintsch, 1998). In the same fashion, existing theories and experimental methodologies of perception and comprehension can be applied to describe and study visualisation comprehension. A coherent account of the processes involved in the visualisation comprehension will then facilitate the assessment of comprehension, and, by extension, quality assessment of visualisations.

This thesis applies the theories and methodologies of cognitive psychology to the quality assessment of novel visualisation techniques, and lays down foundations for a theory of graph comprehension. The resulting account will be tested against a series of concrete tasks, based on evaluating comprehension of a specific visualisation. The aim is to produce novel findings on three levels of applicability. The first level is applied and immediate: to assess one specific

visualisation technique. The second level is methodological: to apply the toolset of cognitive psychology to the evaluation of visualisation techniques. The last level is theoretical: to contribute to the development of a theory of visualisation comprehension.

A focused interest in the study of visualisations is a relatively recent phenomenon, going on for less than half a century (Friendly & Denis, 2001), and so an established terminology is still forming. In the next chapter, to increase clarity, a consistent set of terminology will be set out, providing a working map of the field suitable for further discussion. Thus Chapter 2 starts with a philosophical review on the differences between data and information, provides several useful taxonomies of the field, and continues with a review of theoretical and evaluative studies concerning visualisations.

The empirical part of this thesis begins in Chapter 3 with a preliminary study of visualisation practices used by researchers within the University of Abertay. This preliminary study is designed to establish a specific exploratory visualisation technique and task to be used as the applied example for comprehension evaluation process. Based on the insights gained from the preliminary study, Chapter 4 outlines the specific cognitive theories and methodologies applicable to the research of visualisation comprehension. Chapter 5 summarises the methodology used in the rest of this thesis, including the psychophysical paradigms used, the statistical analyses applied and the hardware and software developed for and used in the main experimental body of the thesis: experiments in visualisation comprehension with human participants. Chapters 6, 7, 8 and 9 detail the motivations, methods, results and outcomes of each of the experiments. Lastly, the overall discussion of the results in terms of the theoretical framework outlined in Chapters 4 and 5 is presented in Chapter 10, which also summarises theoretical and practical implications of this thesis and presents recommendations for further research.

Chapter 2. Visualisation research: a patchwork paradigm

Introduction

This chapter provides a synopsis of the novel and constantly evolving field: the study of visualisation, with a view towards finding a focus area most in need of further attention. At the outset, basic terminology needs to be clarified. Since the focused interest in systematic study of visualisations is relatively recent and comes from many different fields of knowledge, the relevant vocabulary is not wholly consistent. It has been noted, for example, that the term ‘visualisation’ itself “*can refer to a research discipline, to a technology, to a specific technique, or to the visual result*” (Van Wijk, 2005). Research in visualisation comes under related but different titles of Data Visualisation, Scientific Visualisation, Information Visualisation and Visual Analytics. It is therefore important to introduce several key terms in the way that they will be used in this thesis. The main part of this chapter will then provide a review of the visualisation research that offers evaluation frameworks, focusing on studies where the evaluation draws on cognitive psychology.

Terminology

The most complex of the terminological distinctions is the difference between *data* and *information*. Both terms are originally Latin: ‘information’ is derived from *informare*, ‘to change, give form or shape the subject’, and ‘data’ is the plural of *datum*, literally ‘the givens’ (Brown, 1993). Both are abstract, and their definitions and interconnection draw deeply on a philosophical debate regarding the nature of knowledge (e.g. Floridi, 2005). In this thesis, the following definition will be used: *data* are the facts about the world, results of observations, measurements and calculations; *information* is of the domain of the mind, closely related to meaning, knowledge, instruction, communication and understanding (Black, et al., 2009). Data can be stored and represented in many forms: numerical, pictorial, verbal etc. Information, on the other hand, cannot be easily recorded, described or measured,

but it is crucial in informing decisions and actions. According to these definitions, the processes of perception and cognition, crossing the boundary from the physical to the mental, transform data into information. An important observation regarding the connection between data and information is that data yield information about a specific *topic*. Without a topic or task, there is no information, and different reasons for using the same data will result in different information.

The colloquial use of the term ‘information’ is fundamentally different from the one outlined in the previous paragraph. It is derived from Shannon’s law (Shannon & Weaver, 1949) which introduces units of measurement for information: bits and bytes. In fact, this measurement is not of the amount of information, but rather of a theoretical upper limit of information that can be stored in a system with a limited number of states, or, in other words, in a datum of a given size. Shannon’s definition, on one hand, made possible objective externalisation of information in fields of informatics and computer science, but on the other, it gave rise to the current common use of the word ‘information’ which can be approximated to “the stuff you need a computer for”, as evidenced by terms such as ‘Information Technologies’, ‘Information Services’ etc. This interpretation of the term will not be used in the current thesis.

There are currently several research fields concerning themselves exclusively with the study and development of visualisations. These are: Data Visualisation (C. Chen, Härdle, & Unwin, 2008), Information Visualisation (Ware, 2000) and Visual Analytics (Thomas & Cook, 2005). All three deal with the same subject matter: the visual representations of data conveying information to human users. The difference is in the areas of knowledge that provide the tools used in dealing with that subject matter, and in the stages of the process between data, chart and user that are at the focus of each discipline. Data Visualisation focuses on the visual representation of complex data sets, on the conversion from data to the chart, using tools of applied mathematics and computer sciences (C. Chen, et al., 2008). Information Visualisation focuses on design of novel visualisations representing specific

mental models, drawing on tools from computer graphics, but also multimedia art and design (Chaomei Chen, 2004). Information Visualisation mainly deals with the chart itself. Finally, Visual Analytics focuses on user experience given the specific aims for which a given visualisation is designed (Thomas & Cook, 2005). Visual Analytics is close to human factors studies, human-computer interaction research and usability studies.

Of course, studies concerning visualisations can be found in other fields as well, e.g. in pedagogy (see Freedman & Shah, 2002 for an excellent review). Specifically in the field of Human-Computer Interaction (HCI) there are a number of studies involving visualisations (e.g. Coutaz, et al., 1995) but they are mostly concerned with the final usability evaluation of complex tools that are created for a specific task and audience in mind, and not with fundamental visualisation techniques. The works of Wickens and colleagues (e.g. Wickens, Merwin, & Lin, 1994) are a rare exception, bringing the methodological apparatus of HCI to study elementary 3D visualisations such as surface plots and scatter plots; they will be discussed in more detail in the last chapter of this thesis.

The Science of Visualisation

Generally speaking, there are many different ways in which research in one field can be based on findings from another field. A natural science, such as psychology, generates knowledge on four broad levels of abstraction: empirical observations; tools and methodologies; theories; and lastly, laws and models (Kuhn, 1996). Currently the body of knowledge produced by psychology research includes many facts and methodologies, a number of theoretical accounts and a small number of mathematically expressed models and laws, e.g. models of memory (Norman & Bernbach, 1970) or Fitts law (Fitts, 1954). Research of visualisation based on psychology can draw on any of these. There are many non-experimental studies that apply findings generated by visual perception research to generate novel visualisation techniques (e.g. Interrante, 2000; Nowell, Hetzler, & Tanasse,

2001; Ware & Knight, 1995). The reviews in this chapter will instead focus on more general work based on theories of perception and cognition, and on evaluative studies.

Two major works provide an appropriate starting point for a review of visualisation research. One is an influential textbook “Information Visualisation: Perception for Design” (Ware, 2000) which grounds best practices in information visualisation design in empirical findings regarding human visual perception. The other is a result of comprehensive research into the state of the art in the field from the US National Visualization and Analytics Centre (Thomas & Cook, 2005).

In the first chapter of his book, Ware proposes a new ‘science of visualisation’. The book outlines several taxonomies of visualisations and reviews paradigms and research methodologies of related areas such as psychophysics, sociology and human-computer interaction. Importantly, Ware challenges the applicability of the semiotic approach to visualisation design (Bertin, 1967), arguing that, for the most part, the language of visualisations is not a collection of arbitrary symbols. The main bulk of the book provides empirical evidence from vision research and outlines its possible implications for visualisation design. It is organised in chapters according to visual features such as colour or depth, with little generalisation across chapters. Overall, the book defines the focus areas well, but does not present a single coherent paradigm in a Kuhnian sense (Kuhn, 1996) that would qualify as a science of visualisation on its own right.

The major value of Ware’s book is that it provides, in a clear language and open style, a common ground and framework for the development of the ‘science of visualisation’. Ware offers a useful description of the process of visualisation in terms of five main entities: the *real world*; the *data*, produced by measuring some aspects of the world; the *image*, created from this data according to the visualisation mapping function; and the *viewer* who is the user of the process and who has a *task* to perform. The aim of the visualisation process is to inform the *viewer* about the *world* sufficiently for the completion of the *task*. This

description is in agreement with the definition of information given above (Black, et al., 2009), as well as with other accounts of visualisation (e.g. Van Wijk, 2005).

Ware, as well as other authors (e.g. Chaomei Chen, 2004) introduces *models* to describe task-specific instances of data, visualisation and information. *Data model* defines the structure of the data that needs to be analysed, describing the variables that it contains and their properties such as units of measurement, precision, resolution and range. *Visualisation model* is a description, in the same terms, of the representation space. For example, a visualisation model of a scatter-plot includes the vertical and horizontal coordinates of each point, the point size, shape and colour. Lastly, *information model* is the description of the structure of the message that can be inferred from the data; it is related to schemata and mental representations. Information models are constructed in the mind of the user, and are therefore less straightforward to describe.

The second major work, “Illuminating the Path” by Thomas and Cook (2005) is unique in that it has a clear and direct applied agenda: aiding homeland security, i.e. the defence effort against terrorist activity aimed at the US territory, citizens or property. The advances in Visual Analytics are reviewed against that agenda, resulting in a book that is singularly grounded in the real world. The authors identify four relevant areas of research: the science of analytical reasoning; visual representations and interactions; data representations and transformations; and production, presentation and dissemination of information. The book includes comprehensive reviews of each of these fields, as well as recommendations for future research in each of these areas. One of these recommendations is to “*Create a science of visual representation based on cognitive and perceptual principles that can be deployed through engineered, reusable components*” (ibid., p. 8), another is to “*Develop an infrastructure to facilitate evaluation of new Visual Analytics technologies*” (ibid., p. 14).

The recommendation of “*creating a science of visual representation*” is somewhat odd in the year 2005, given that, according to Ware (2000), such a science already exists. In fact, the scientific foundations for the study of graphical methods for data analysis were discussed much earlier. Cleveland and McGill (1984) write: “*The subject of graphical methods for data analysis and for data presentation needs scientific foundation*” and follow by providing such foundations. Their account is based on the notion of ‘elementary perceptual tasks’, e.g. length comparison, area comparison, hue comparison etc. Every visualisation technique is thus ranked in terms of the elementary perceptual tasks that it employs, and how easy they are; the unspoken assumption being that these tasks can be uniquely and linearly ordered by the ease of perception. Several years after that, Csinger, in his “Psychology of Visualisation” (1992) challenges this assumption, laying down a model incorporating theories of perception and psychophysical methodology resulting in the automated generation of a visualisation technique tailored for a given data model. Van Wijk (2005) proposes a mathematical flow model of data to image to perception to knowledge, with value-functions estimating efficiency and effectiveness of each step. Most recently, Green, Ribarsky and Fisher (2009) refine and extend this model into a Human Cognition Model (HCM), incorporating processes such as creation and analysis of hypotheses based on visual evidence.

A common foundation of any natural science is a taxonomy of observed phenomena (Atran, 1993). Several taxonomies of visualisation techniques have been proposed. Some are based on the properties of the data model, e.g. whether the data are discrete or continuous; or the extent to which data semantics predefine and constrain the visualisation (Tory & Moller, 2004). Other bases for classification have been used as well: a taxonomy based on multisensory visualisation model options (Nesbitt, 2004), a taxonomy specifically for classification of different types of ‘dirty’ data input (Li, Peng, & Kennedy, 2010), etc.

To summarise, the study of visualisations is a vibrant research area, drawing from many diverse fields. The task of creating a ‘science of visualisation’ is well underway, tackled from many different directions and using various neighbouring disciplines: computer

science, especially computer graphics and data mining; visual design; statistics and areas of mathematics such as graph theory; psychology of perception and cognition; education; economics; and many more. However, as much as it is necessary and fruitful, this diversity is at times counter-effective, since the ‘parent’ paradigms differ greatly in the understanding of what constitutes an area of research, what can be a valid methodology, and what evidence amounts to a proof.

For instance, only a small number of the abovementioned theories and taxonomies account for factors external to the user and the tool, such as goals, tasks and actions resulting from the interaction with the visualisation. Even the researchers that stress the role of task and context in the process of visualisation use (e.g. Van Wijk, 2005) do not include them in their models. This is despite ample evidence that goals modify perception and decision making on all levels (Blake & Sekuler, 2005). For example, an interview-based research of journalist practices (Attfield, Blandford, Dowell, & Cairns, 2008) stresses the importance of ‘angle’ (essentially the task given to a journalist) for visual data analysis. Another piece of anecdotal evidence can be seen in a number of visualisation tools, the success of which can be traced to a design process that is well-grounded in domain expertise (e.g. Graham, Kennedy, & Benyon, 2000).

Moreover, none of the reviewed theories include a falsifiable predictive element, e.g. an evaluation methodology that is based in the theory. The authors of HCM (Green, et al., 2009) are aware of that need, and plan to develop their model to have predictive power in the future. In the works that employ theories of perception and cognition (e.g. Ware, 2000), there is little attempt to trace visualisation processing beyond low-level perception, to high-level processes such as the construction of insights, information and knowledge – which are the main goal of using visualisations. One notable exception (Freedman & Shah, 2002) incorporates pre-existing knowledge (but not task constraints) and produces testable predictions. Unfortunately, it went virtually unnoticed by the visualisation research community (29 citations recorded by Google Scholar at the end of 2010).

As a result, there is a discrepancy between the state of the science of visualisation as reviewed above and the science of visualisation as envisioned by the authors of “Illuminating the Path” (Thomas & Cook, 2005). “Illuminating the Path” is highly goal-oriented, demanding verification against real-world cases as well as a comprehensive theoretical foundation. The need for a theoretical account of the focal concept in visualisation research, that of insight, as well as for a testable evaluative framework of assessing these insights, has not been fully addressed to date.

It is possible to take one step further and wonder whether there is a science of visualisation at all: one of the main criteria for being a science is having a theoretical account generating falsifiable hypotheses (Popper, 2002); another is a single internally coherent research paradigm grounded in empirical evidence (Kuhn, 1996). The accounts reviewed above generally fall short of having a single theoretical account focusing on the major components of the researched phenomenon, and from which falsifiable hypotheses can be generated.

To meet these criteria, a ‘science of visualisation’ would first of all have to have its own terminology which will describe entities that are central to visualisation design and usage. This terminology could arise from existing taxonomies, and the central entities that this science would deal with would not be borrowed from them other fields, e.g. mathematics, cognitive science or visual design – even though they could be reducible to (or at least explained by) entities in these disciplines. The ‘science’ would develop metrics and unique research methodologies focusing on these entities. That would result in a paradigm that is internally valuable, coherent, and verifiable against empirical evidence; a ‘normal science’ according to Kuhn (ibid.).

Summarising the benefits and shortcomings of the reviewed research, Figure 2 presents a collation of the mind-maps and process diagrams that are found in the works reviewed above (Csinger, 1992; Tory & Moller, 2004; Van Wijk, 2005; Ware, 2000), with the addition of missing elements that are highlighted by Thomas and Cook (2005). Three concentric

feedback loops form the general structure of this model. The outer loop is the interaction of the user with the world, tying the whole process to the real world by a *Task* assigned to analysts and the *Actions* that they perform as the outcome of the visual analysis, and also by the *Measurements* that introduce *Data* into the system. The second loop is the interaction with the visualisation, with the user adjusting the parameters of the *Visualisation* and thus changing the *Display*. The final loop is an abstraction of the processes of guided attention and is fully internal to the user; it is shown as iterative fine-tuning of *Perception* and *Decision-making* by the *Context*.

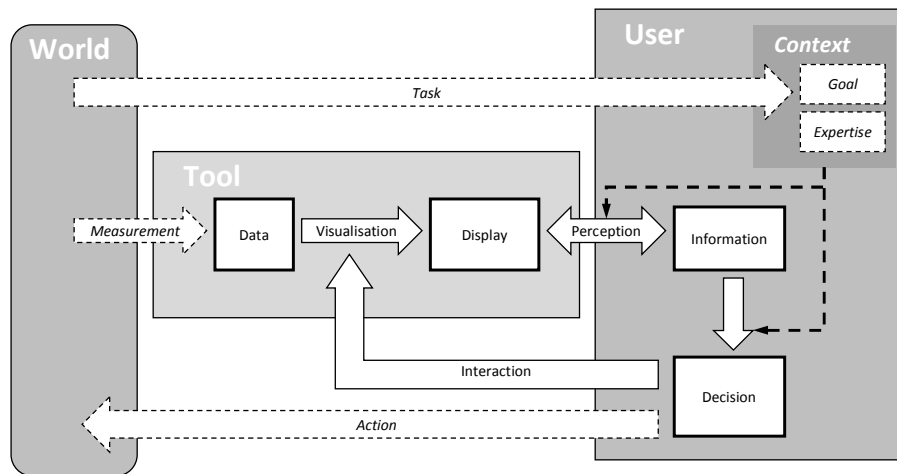


Figure 2 – a model of processes involved in visualisation comprehension. Elements that were added by the author and are not commonly present in literature are marked by dashed lines and italics.

The lack of commonly accepted unifying paradigm affects the practice as well. While evaluative studies of visualisations abound, a recent meta-analysis (Amende, 2010) shows that in many of them quality evaluation plays only a token part in the description of a novel visualisation developed by authors. This inevitably affects the scope and depth of such studies and, unfortunately, can sometimes cast doubts over their objectivity. The meta-analysis attributes that practice to an absence of theoretical models of visualisation success. Likewise, a review of different approaches to evaluating the effectiveness of visualisations (Zhu, 2007) concludes that there is no accepted definition for the effectiveness of visualisation comprehension. This lack of effectiveness criteria for visualisation

comprehension is an inevitable result of the need for a model for visualisation comprehension itself.

Even the best examples of evaluative research suffer from that lack of theoretical framework. For example, in a series of experiments (Arthur, Booth, & Ware, 1993; Ware, Arthur, & Booth, 1993) investigating the relative benefits of stereo rendering and head-tracked camera control for abstract data visualisation, accuracy and speed are measured in an abstract 3D task of tree-tracing (see Figure 3). The results indicate roughly additive improvement in accuracy for stereo presentation and head coupling. However, it is not clear how much these results can be generalised to other visualisations and tasks using the same rendering techniques. Originally (Sollenberger & Milgram, 1993), the tree-tracing task was developed as an emulation of a real-world domain-specific task: tracing blood vessels in a brain scan. However, without an account of processes involved in perception and comprehension, there is no way to determine the extent to which this task is a representative of abstract 3D visualisation tasks, or even what is the similarity between it and another task such as, for example, recalling the structure of a 3D surface (Wickens, et al., 1994).

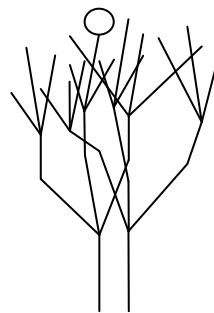


Figure 3 – a schematic stimulus from the tree-tracing task. In a typical task, two intertwined 3D structures are presented in 2D, stereo, or head-tracked movement, and participants are asked to identify the structure to which the highlighted leaf belongs.

Summary

In current studies of visualisation, diverse research areas address visualisations from different perspectives and in conjunction with other disciplines, but there is no single ‘science of visualisation’. There are many attempts to formulate a science of visualisation,

but most fail to account for the core question of how perceiving and interacting with charts produces insights, generates novel information, and aids in problem solving and decision making. There are many excellent evaluative studies, but due to the lack of the theory of visualisation comprehension, these studies assess and compare measurables that are difficult to generalise, and that are open to criticism of internal and external validity.

To provide a theory of visualisation comprehension, one of the first questions that has to be answered is: what constitutes visualisation comprehension? What kind of information is yielded by visualisations, or, in other words, what types of insights can be gained by using visual analysis? These questions can be addressed directly, by an empirical study of the visualisation users. Such evidence would be valuable since, as seen in the above literature review, there is currently a relative scarcity of structured studies of the general end-user practices. Therefore, a study of current visualisation practices is described in the next chapter.

Chapter 3. Preliminary study: visualisation practices in research

Introduction

One of the conclusions from the literature review in Chapter 2 is that in Visual Analytics there is a scarcity of rigorous, structured studies of end-user needs, requirements and practices. One of the reasons that makes the studies of current practice so challenging to undertake is that both needs and tools are constantly changing; needs change because of the progress in research; and tools change due to new developments in both hardware and software. Therefore, a structured study of the end-users' requirements was planned as the starting point of the current research into visualisation efficiency. A group of researchers in diverse fields from the SIMBIOS/UTWC¹ multidisciplinary research centre kindly agreed to participate in this study.

A range of techniques is available for such a study (Langdridge & Hagger-Johnson, 2009), from free interviewing to rigid multiple-choice questionnaires. A completely unstructured interview, a method more suitable for autobiographic and ethnographic enquiry, may easily lose relevance to the topic explored. A completely structured interview or a closed-question questionnaire would, on the other hand, unnecessarily constrain the interviewee and may miss important points because of the experimenter's preconceptions and biases. For this study, a semi-structured interview provides a necessary balance between the competing constraints of open-endedness and lack of bias on one hand and relevancy and ease of analysis on the other. The results given by a semi-structured interview can be expected to be relevant, ecologically valid and without much bias (*ibid.*).

¹ SIMBIOS is the Scottish Informatics, Modeling and Biological Systems centre; UTWC is the Urban Water Technology Centre, both in the University of Abertay, Dundee.

Formerly, qualitative research methods in psychology, including the open-ended semi-structured interviews, used to be considered one of the less scientifically rigorous tools available to the psychologist. Currently, qualitative methods are increasingly seen as a complementary tool to quantitative methods, and are used for research into subjective opinions and for hypotheses generation. Experts in qualitative methods have successfully demonstrated that, with due care, scientific rigour can be achieved with this tool (Langdridge & Hagger-Johnson, 2009). One example of measures that can be taken to ensure better validity is to ask questions about a specific experience, rather than general or hypothetical questions. It is preferable to let the participant choose the experience, as long as it is relevant to the question, i.e., in our case, includes information visualisation as a major stage of data analysis.

The focus of the current survey is data visualisation practices. To account for personal differences, this is put into context of individual researchers' professional background. Details of the types of data that are analysed and the research predictions and findings are taken into account. The aim is to disclose current practices with the view of constricting further research to the most relevant areas of the field.

Method

Every interview started with a brief introduction and explanation. After the interviewee has signed an informed consent form (see the enclosed CD) the audio recording of the interview would start. The questions were asked by the interviewer and discussed between the interviewer and the interviewee until the interviewee felt that the question had been sufficiently addressed. The questions were as follows:

- A. What is the general area of your research?
- B. Please tell me about a piece of research you think would be a good example of the data analysis process you generally use.
 - a. What was the research question?

- b. What data did the research produce?
 - i. Data type
 - ii. Data dimensionality and size
 - iii. Whether the data has built-in spatial or temporal qualities
 - iv. Other characteristics
 - c. What were the stages in your data exploration (4-5 stages)
 - d. How did you visualise the data during these stages? Could you sketch it?
 - e. What tools (computerised or manual) did you use to visualise it?
- C. Could you tell me about your research background and current affiliation?
- a. What is your academic background?
 - b. What would you define as your areas of expertise?
 - c. What is your current affiliation?

The responses were not timed, and the interviewees were encouraged to provide as much detail as they liked as the result, interviews lasted between 30 minutes and an hour. Pen and paper were provided for sketches of data structure and visualisation designs. In some cases, the interviewees offered to demonstrate the data, visualisations and analyses on their computers. In these cases, the sketches were done by the experimenter.

The interviews were recorded using a portable digital voice recorder, and the recordings were later manually coded and analysed in conjunction with the sketches. The analysis was for recurring patterns in either of the answers or sketches.

Results

Five researchers (4m/1f) agreed to be interviewed. Their research was in the areas of plant genetics, river hydrology, forensic psychology, geology and physics. Data that they produced and analysed in their research varied across the entire spectrum of data taxonomy: categorical and interval, subjective ratings and precise objective measurements, with and without inherent constraints on visualisation. For data visualisation, some used relatively

common tools such as Microsoft Excel, MATLAB or SPSS; others used highly specialised software solutions for volumetric or geographical data. One visualisation format, a two-dimensional scatterplot, was mentioned by four out of five interviewees. Other common formats, such as histograms, bar-charts, pie-charts etc. were also mentioned, although less frequently.

The recurring theme that appeared in every interview was the intended use of visualisation. When asked about tasks that were solved by visualising/plotting data, all the interviewees described a situation where plotting data helped them to identify outliers in the data. Some of these outliers were due to measurement errors and had to be taken out; others were of genuine scientific interest. In both cases, however, the insight gained from visualising data was the same: outlier detection. Two other tasks had appeared in several interviews, although with lesser frequency than outlier detection. Some interviewees (3/5) reported using data visualisations to find whether the data can be divided into distinctive sub-groups or clusters, with these clusters then becoming the focus for later analyses. Others (3/5) reported focusing on the dynamics, shapes, patterns and trends, in order to reveal relationships between data variables.

Discussion and Conclusions

To summarise, the recurring theme in the interviews was that the data are visualised in order to detect three types of entities: outliers, clusters and trends. In all other aspects, such as area of research, types of data, visualisation tool and techniques, there was no consistency of approach between the interviewees.

The definitions of the terms ‘outlier, cluster and trend’ in this work are consistent with the common use in the data analysis literature (e.g. Du Toit, Steyn, & Stumpf, 1986; Tabachnik & Fidell, 2006). Specifically, clustering is a separation of data into a small number of discrete subsets, or *cluster*; when there is no such separation, the data can be said to form a single cluster. *Trends* are patterns or regularities of the data in a cluster. Finally, *outliers* are

a relatively small number of data points that clearly do not belong to any of the clusters and do not conform to any of the trends (see Figure 4 for an illustration of 1D and 2D cases). Essentially, the detection of these entities in a visualisation constitutes a parsing of the visual stimulus into meaningful entities, the existence of which is later confirmed by relevant statistical analyses. There is a functional difference between these three entities as well: outliers are usually discarded, while the separation of the data into clusters and subsequent analysis of trends within clusters constitute the main body of the data analysis.

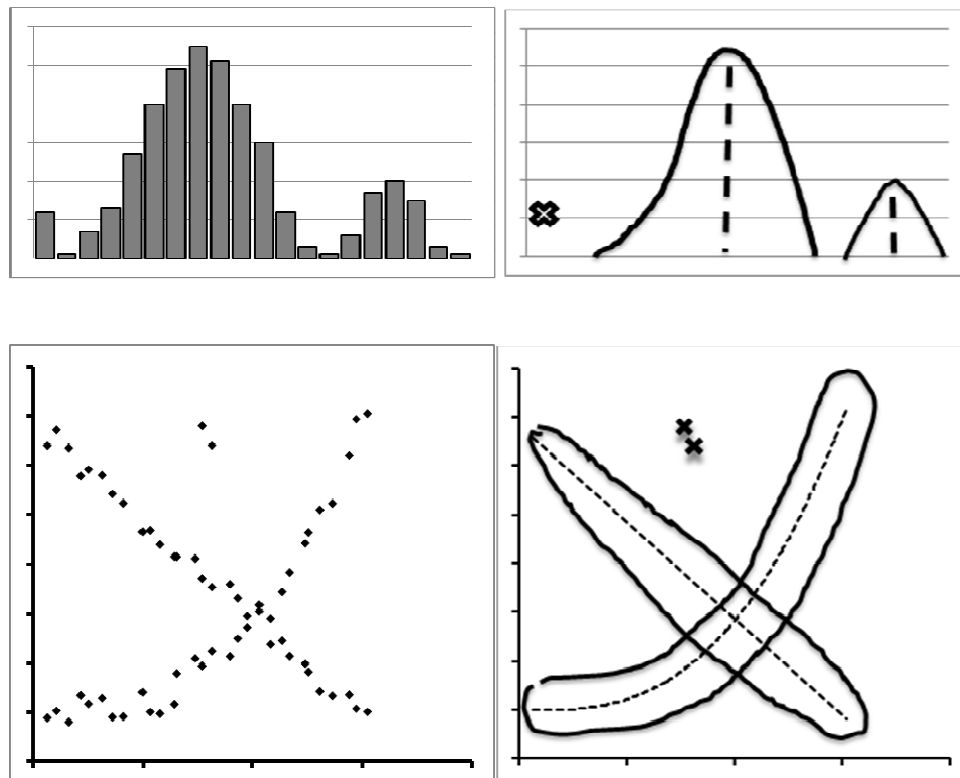


Figure 4 – outliers, clusters and trends in 1D and 2D visualisation. Top: 1D histogram; bottom: 2D scatterplot. Left: the visualisation; right: schematic representation of parsing into outliers, clusters and trends. Outliers are denoted by X, clusters by solid lines and trends by dashed lines.

From a methodological point of view, it can be seen that the open-ended semi-structured interviews produced results that were consistent, relevant to the question asked and yet, unexpectedly, not directly related to any of these questions. A more structured interview may not have identified the commonality in the purposes of visualising data. Similar research that concentrated on tools and techniques used, but not on tasks performed and insights gained, did not find any uniformity between researchers (Robinson, 2009).

It is easy to see why trend detection, specifically, is better addressed with the aid of visualisations. Non-visual means, such as correlation or regression analysis allow detection of linear trends only. Through the medium of data visualisation, non-linear trends, such as polynomial, repetitive or exponential, can be spotted with the same ease as linear trends (Anscombe, 1973). Detecting non-linear relationships in the data without visualisation requires complex statistical analyses, different for each type of relationship. This difference is especially important when no prior knowledge of a relationship within the data exists, and the analysis is exploratory, a part of the hypothesis-generation process. A similar argument can be made for outlier and cluster detection: these tasks are performed by visualising data because this way these structures can usually be found rapidly and effortlessly. Research methods literature (Tabachnik & Fidell, 2006; Tukey, 1977) suggests using visualisations as the easiest method to detect outliers. This is also what was found to be the common practice of the sampled University of Abertay research community.

Chapter 4. Towards a theory of visualisation comprehension

“Graphs are a recent invention, and if they are an especially effective method of communication, it is because they exploit general cognitive and perceptual mechanisms effectively” (Pinker, 1990)

A theoretical framework that seems well-suited to studying the three tasks that emerged from the preliminary review of existing practices is a combination of the Principles of Perceptual Organisation (Kubovy & Pomerantz, 1981) and the Reverse Hierarchy Theory (Hochstein & Ahissar, 2002). This chapter outlines these two theories and justifies their relevance to the current study.

Principles of Perceptual Organisation describe the cognitive processes by which a visual sensation is organised and parsed into meaningful objects. Recently it has been noted that these processes are not exclusive to the visual system, with similar principles being proposed for the formation of auditory (Bregman, 1994) and haptic (Chang, Nesbitt, & Wilkins, 2007a, 2007b) objects. Multimodal studies have extended the conceptualization of perceptual organization beyond the familiar modality-bound accounts deriving from Gestalt psychology leading, ultimately, to principles of multimodal perceptual organisation (Remez, Fellowes, Pisoni, Goh, & Rubin, 1998). The existence of these processes and a first formulation of principles governing them were posited almost a century ago, as the underlying principles of Gestalt theory of perception: the Laws of Gestalt (Wertheimer, 1938).

The Laws of Gestalt have gained recognition in areas beyond psychology, becoming, in effect, a meme². On one hand, every once in a while they are declared outdated and irrelevant (e.g. in Marr, 1982), but on the other, they are still very much a part of discourse and

² *meme* is defined as a postulated unit or element of cultural ideas, symbols or practices (Dawkins, 2006)

curriculum of psychology (Blake & Sekuler, 2005), visual arts (Zakia, 1975), human-computer interface design (Dix, Finlay, Abowd, & Beale, 2004) and information visualisation (Ware, 2000). The reason for their popularity is clear: they are powerful, they are robust, and they produce the “eureka effect”. Just looking at any one of the popular gestalt illustrations, such as Necker’s cube (Necker, 1832), Kanizsa’s pac-mans (Kanizsa, 1976), Rubin’s face/vase (Rubin, 1921) and many others (see Figure 5) makes the observers feel that they begin to understand how their own mind works.

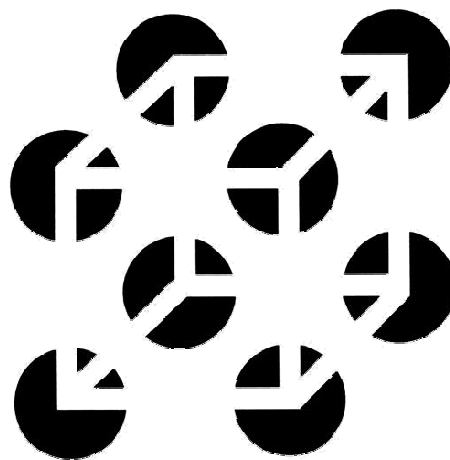


Figure 5 – optical illusion combining effects of grouping, figure-ground segregation, and illusory contours to create a perception of three-dimensional white cube from sensation of 24 black sector shapes.

However, exactly because of this popularity they are often subject to misinterpretation or oversimplification. The strength of the visual experience, paradoxically, can be a source of the problem. The visual illustrations are often so striking, and the basic entities such as objects and elements so intuitively obvious, that not many feel the need to objectively define the terminology or measure the effect. Those that do, find it to be non-trivial (Kubovy & Pomerantz, 1981; Scholl, 2001), leading to fundamental philosophical dilemmas on ontology and epistemology. On a tangential note, a similar problem has been noted regarding visual proofs in mathematics: appealing visual illustrations sometimes tend to hide the shortcuts and even fallacies in the logic of the proof (Casselman, 2000).

The next paragraphs outline the gestalt theory of perception, from its basic tenets established in the early 20th century to the latest findings, focusing on objective criteria and methodologies. In order to stress the more problematic aspects of this theory that are often overlooked due to the visual strength of the illustrations, the description is purely textual, with no illustration presented. The popular view of the gestalt laws and the gestalt school of psychotherapy (Perls, Hefferline, & Goodman, 1951) will be left out of the scope of this review.

In the late 19th century, the study of perception was dominated by a structuralist school, with the agenda of cataloguing the atomic elements of visual stimuli, such as hue, lightness etc. (e.g. Blake & Sekuler, 2005). The main novelty of the gestalt approach in 1910s was the shift of the focus from these basic elements to the ways they may be combined into objects. The objects, the theory argued, while visually composed of the basic elements, are perceptually different from the sum of their parts. The particular arrangements of elements that result in these objects were called *Gestalts*. Gestalt was closely linked to a philosophical notion of emergence, “associated with features such as novelty, irreducibility, and unpredictability...novel properties that are neither predictable nor explainable on the basis of ...the parts” (Achim Stephan, 2003, as cited in Pomerantz, 2006).

The conditions that describe which arrangements of elements become a gestalt are commonly known as the Laws of Gestalt. The underlying assumption was that there are only a limited number of these conditions, and it is therefore advantageous to study them. This assumption was challenged by some studies where laws of gestalt number in hundreds (Chang & Nesbitt, 2006). Still, most of the laws formulated to date describe only four cognitive processes, which are, moreover, related: grouping; figure-ground segregation; the principle of prägnanz³; and reification (Blake & Sekuler, 2005). These are elaborated below.

³ prägnanz (German) – pithiness

Laws of *grouping* describe the conditions under which several elements of the visual stimulus are perceived as part of a group, that can be perceived either as an object, or, as some researchers emphasize, as a texture (Julesz, 1981). Many laws of grouping have been described, among them grouping by spatial proximity, grouping by similarity (for example by shape, colour or luminosity), grouping by common fate (common movement, or proximity in space and time) and grouping by closure (proximity in space, gradient and curvature).

Figure-ground segregation is a process related to grouping. Whenever grouping divides the visual field into two regions, one is then perceived as the object and the other fades to, and becomes part of, the background. An important feature of figure-ground segregation is its *bistability*: in many (but not all) groupings each of the two parts can be perceived as the object, thus making the other part into a background. This swapping of the perceived object and background can be performed both voluntarily, by directing attention, and involuntarily (Driver, Davis, Russell, Turatto, & Freeman, 2001).

The *principle of prägnanz* limits the bistability by stating that the process of perceptual organisation will tend to converge on the most succinct of the possible sets of objects that may be the representation of a given perceptual stimulus. This principle is often called the Occam's Razor of perceptual organisation (Kubovy & Pomerantz, 1981).

Lastly, *reification*, also known as the formation of the illusory contours, is the process by which the perceptual system assumes the existence of elements that are not present in the sensorium, if they allow for a better representation of the scene according to the principle of *prägnanz*. Such illusory elements are then treated by the processes of grouping and figure-ground segregation in exactly the same manner as the elements that are present in the perceived stimulus, allowing for a complex multi-tiered parsing of visual sensation into meaningful objects. In fact, in several famous optical illusions, such as Kanizsa's triangle (Kanizsa, 1976) or the mega-phi effect (Steinman, Pizlo, & Pizlo, 2000), all the elements that

are part of the visual sensation become part of background, and the perceived object is composed solely of illusory elements.

One of the problems with the Laws of Gestalt was the introspective methodology used by Wertheimer et al. in their formulation (Blake & Sekuler, 2005). Introspection has been heavily criticised as being subjective and non-generalisable. This resulted in, among other problems, overabundance of ‘laws of gestalt’; some authors count as many as 700 (Chang & Nesbitt, 2006). The introspective method is now generally dismissed in psychology except as an early stage of hypothesis generation, and more objective methodologies, such as psychophysics or neurophysiology, have replaced introspection as the main tool of psychological research. Some of these have been successfully applied to the study of the Principles of Perceptual Organisation. For example, the neuropsychological research has led to discovery of the neural correlates of the ‘gestalt moment’ (Keil, Muller, Ray, Gruber, & Elbert, 1999).

Recently, psychophysical methods were proven to be useful in establishing measurable, objective and verifiable criteria of ‘gestaltness’, i.e. which assemblages of elements constitute an object and which do not (Kubovy & Pomerantz, 1981). The main method used is *mental chronometry* (Donders, 1969), i.e. measuring the reaction time of human participants and comparing average reaction times in different conditions. According to (Pomerantz, 2006), gestalt is an arrangement of elements that demonstrates ‘*pop-out*’ effect, ‘*pop-out*’ asymmetry and *configural superiority*.

The ‘*pop-out*’ effect was first observed when measuring participants’ reaction time in a task that required them to identify an odd-one-out in an array of otherwise identical elements (Treisman & Gelade, 1980). Two distinct patterns of results were observed: in some cases, manipulating the number of non-target (distracter) elements in the stimulus array led to a linear increase in average reaction times; in other cases, identifying the odd-one-out took

constant time regardless of the number of distracter elements presented. This latter phenomenon has become known as the ‘pop-out’ effect.

‘Pop-out’ asymmetry is a ‘gestaltness’ criterion that is secondary to the ‘pop-out’ effect. It is tested by switching the target and the distracter element, i.e., looking for the previously non-target element in an array of what was previously the single target. This change usually retains the ‘pop-out’ effect, but leads to different reaction times. Gestalt, then, is defined as the configuration of elements that is identified faster when it is the ‘figure’, rather than when it is the ‘ground’ (Ahissar & Hochstein, 1996; Wolfe, 2001).

The *configural superiority effect* is observed when the detection of an element as a part of a gestalt is much faster than the detection of the same element on its own (Pomerantz, Sager, & Stoever, 1977). In their typical experiment, participants are presented with a display containing four items in a 2x2 matrix and asked to indicate the item that differs from the other three. In some conditions, adding visual elements carrying no task-relevant information was found to improve performance. It has been hypothesised that perception is facilitated by novel properties such as closure that emerge when the additional elements combine distinct features into a unified configuration.

From the above description, it can be seen that the ‘pop-out’ effect is commonly used as a testing methodology in the research of the processes of perceptual organisation. The general distinction between two modes of perception, the immediate ‘pop-out’ and one-by-one serial search, is robust, and has been demonstrated in a range of setups and conditions (see Blake & Sekuler, 2005). The question that remains is: why this effect is a valid psychophysical indicator for gestalt? What cognitive processes occur during the performance of an odd-one-out task, and how do differences in observed outcomes between immediate and serial perception take place? Since its discovery in 1980, several theoretical accounts for the cognitive processes involved have been proposed.

The earliest theoretical account is Feature Integration Theory (Treisman & Gelade, 1980), which posits that different kinds of attention are responsible for binding different features into consciously experienced wholes. The theory has been one of the most influential psychological models of human visual attention (Blake & Sekuler, 2005). According to it, in a first step to visual processing, several primary visual features are processed and represented with separate feature maps. These are later integrated in a saliency map that can be accessed in order to direct attention to the most conspicuous areas. A search for a target defined by a single distinguishing primitive feature can be performed fast and pre-attentively. The search for targets defined by a conjunction of primitive features is much slower and requires conscious attention. Thus, Feature Integration Theory associates the observed ‘pop-out’ effect with process of single feature search, and observed serial search with the search for feature conjunctions. A conclusion from many experiments is that colour, orientation, and intensity are primitive features for which feature search can be performed.

Since the early 1980s there has been a growing body of evidence incompatible with the Feature Integration Theory. Mostly it is comprised of experiments showing ‘pop-out’ for objects differing in high-level⁴ semantics, rather than low-level primitive features. It has been shown that faces ‘pop out’ from buildings or tools (Hershler & Hochstein, 2005; Tovee, 1998), and that for very similar arrangements of ‘primitive features’, some exhibit ‘pop-out’, while others do not (Kleffner & Ramachandran, 1992; Pomerantz, 2006). Also, Feature Integration Theory does not provide an explanation for the configural superiority effect; on the contrary, it would probably predict an opposite effect since the addition of features is a prerequisite to ‘feature conjunction’ and thus serial search.

Recently, a theory has been proposed that provides a more coherent explanation for the available evidence. Where Feature Integration Theory connects immediate perception with low-level visual processing, Reverse Hierarchy Theory (Hochstein & Ahissar, 2002)

⁴ In cognitive psychology, the sensory organs are conventionally referred to as the ‘bottom’ of the hierarchy of mental processes and consciousness and reasoning as the ‘top’.

dissociates between fast conscious perception and low-level vision. Contrary to the Feature Integration Theory, the ‘pop-out’ effect is attributed instead to the processes taking place in high-level areas, where large receptive fields underlie spread attention detecting categorical differences. The search for conjunctions or fine discriminations, on the other hand, is assumed to depend on re-entry to low-level specific receptive fields using mechanisms of serial focused attention.

Reverse Hierarchy Theory proposes that perception occurs in two distinct stages. First, the ‘bottom-up’ perception, leading to increasingly complex representations, is automatic and implicit. Then the conscious task-driven decision regarding the attention orientation originates at the hierarchy's top and gradually drives attention downward as needed. Thus, our initial conscious percept matches a high-level, generalized, categorical scene interpretation. For later vision with scrutiny, attention is focused at specific, active, low-level units, incorporating into conscious perception detailed information available there.

Defining objecthood not in physical terms, but through abstract high-level mental representations coincides with the views long held by the philosophy of mind. Already Kant, in his *Critique of Pure Reason* (Kant & Meiklejohn, 1882) considers objects to be mental constructs, bearing little relation to the physical world. The physical world, in Kant's view, was composed of matter. Separation of matter into distinct objects that are in turn composed of elements occurs only in the mind of the perceiver and is highly dependent on mental state (in contemporary terms, task and attention), culture and language. Moreover, Kant's proposed laws that govern perception, transcendental aesthetics, bear close resemblance to, and are a probable early inspiration for the Laws of Gestalt⁵.

Both the Laws of Gestalt and the Feature Integration Theory are widely mentioned in Visual Analytics literature (e.g. Ware, 2000). Principles of Perceptual Organisation and the Reverse

⁵ Author thanks Mr. Simon Fokt for suggesting this link in personal communication.

Hierarchy Theory, on the other hand, are less well-known outside of the perception research community. This discrepancy is somewhat understandable, since the latter two are recent theories that have not been yet incorporated into the discourse outside psychology. However, their relevance to the Visual Analytics agenda warrants more interest (Shovman, Szymkowiak, Bown, & Scott-Brown, 2009).

In the case of Visual Analytics, stimuli, i.e., visualisations or perceptualisations, are complex and abstract, built with the aim of facilitating high-level understanding of the meaning of the scene. Both Principles of Perceptual Organisation and Reverse Hierarchy Theory focus on the cognitive processes that connect low-level perception to the formation of high-level abstract semantics. In addition, these theories offer objective methodologies and measurable criteria for assessing the processes of perceptual organisation. In this thesis, theories and methodologies of Principles of Perceptual Organisation and the Reverse Hierarchy Theory will be applied to the cognitive processes involved in visualisation perception, with the focus on formation of high-level semantics.

Chapter 5. Methodology for visualisation evaluation

Applying Theories of Perception to Visualisation Evaluation

Both the Principles of Perceptual Organisation and the Reverse Hierarchy Theory view perception as a process resulting in the formation of high-level abstract understanding of the sensory stimulus. For natural scenes there exists a certain consensus as to what constitutes their semantic content, and what kind of objects would emerge from the parsing of these scenes: the task-relevant mental representation of physical bodies, including affordances, navigation cues, etc. (Gibson, 1986). In contrast, for abstract visualisations even as basic as bar-charts or scatterplots, the question of constituent objects and top-level semantics has never, to the best knowledge of the author, been addressed.

Defining the top-level semantics of an abstract visualisation is not trivial. According to the theories outlined in the previous chapter, perception and cognition are two parts of a continuous process of iterative formation of meaning (objects, mental constructs) from sensation. This process is constrained, increasingly so for the more abstract categories, by intention (task) and preconceptions (memory, skills, training, etc.). For example, in the odd-one-out detection task, the top-level parsing would be the separation of an object defined by some ‘oddness’ from the background of the rest of the objects.

From the preliminary study discussed in Chapter 3, it is reasonable to assume that for abstract visualisations the resulting abstract mental constructs are those reported by the participants: i.e., outliers, clusters and trends. A task-relevant parsing of the visualisation stimulus into clusters, outliers and trends constitutes the result of the perception and cognition process, i.e., the meaning of the visualisation. It is important to note that the mental construct is not defined in terms of the perceived stimulus, the symbol, but in terms of the signified – the information contained in the data. In other words, we assume that

parsing into clusters, outliers and trends constitutes the information about the data as yielded by a visualisation.

This assumption can be combined with the distinction made by Reverse Hierarchy Theory between fast implicit ‘bottom-up’ and slow explicit ‘top-down’ formation of semantic parsing of the scene. Thus in a chart, as well as in any other stimulus, task-relevant objects would be either immediately obvious – or would require serial visual search to identify. Here, we propose that this observation regarding different modes of perception may be ‘turned around’ and used instead to assess chart quality, distinguishing between efficient and inefficient ways to visualise data. According to this quality criterion, an efficient visualisation would be one that allows implicit perception of required information; in an inefficient visualisation, the information has to be searched for, explicitly and laboriously. In other words, a good visualisation technique would be one in which the meaning ‘pops out’; using a poor technique, meaning would only be found after extended scrutiny.

A key benefit of this quality criterion is that it comes with its own evaluation methodology that is objective, repeatable and non-comparative. This methodology consists of measuring reaction response time in an odd-one-out detection task while manipulating the size of the data set. A constant response time with increasing set size, i.e., the ‘pop-out’ effect, indicates immediate perception and so an efficient visualisation. Response times that are linearly increasing with data set size indicate underlying processes of visual search and, therefore, an inefficient visualisation. Any other pattern, such as linearly decreasing or non-linear, would indicate a lack of validity of this methodology to the task.

To summarise, previous paragraphs propose a connection from ‘pop-out’ testing methodology to visualisation efficiency evaluation. That connection, formed from a chain of conjectures, assumptions and hypotheses logically follows from well-established theories as well as from our preliminary study. However it is still a conjecture, an application of perception theories in a novel field: Visual Analytics. Thus, in order to support it, more specific empirical evidence

is required. Some of this evidence is explored and provided in the subsequent chapters of this thesis.

Validating the applicability of ‘pop-out’ effect methodology to visualisation quality assessment has to be done on an example of some representative visualisation technique and some representative Visual Analytics task. Common Visual Analytics tasks that were identified during the preliminary study were the outlier, cluster and trend detection. Of these, outlier detection is the most frequently mentioned, and it also is exactly the same task as the odd-one-out detection employed in the ‘pop-out’ methodology. The inversion of the criterion target from assessing perception to assessing visualisation quality is straightforward – an effortless, implicit outlier detection, in constant time, is preferable to a laborious search that takes longer for more elements, especially in visualisations where the number of elements can go well beyond hundreds.

The choice of a specific visualisation technique to examine is non-trivial given the available plethora of novel visualisations. However, given the novelty of linking ‘pop-out’ effect methodology to visualisation quality assessment, virtually any choice would contribute to the field. The general criteria for choice were as follows: first, the technique should be common enough, so that the findings would be applicable in a wide context; it should be relatively easy to implement, so that software development time does not become a dominant feature of the research programme. Additionally, in order to strike the balance between novelty and availability, it should be technically sophisticated enough so as to be impossible to do on paper, but simple enough to run on a standard office desktop computer.

An interactive 3D scatterplot meets all these criteria; also Ware (2000) notes that there is currently “*not enough work on 3D scatterplots*”. It is a direct extension of a 2D scatterplot, which is one of the most basic visualisation techniques and is frequently used in research, so presumably all of the intended users would be familiar with the concept. It requires real-time rendering and interactivity, but nothing exceeding the capabilities of a standard desktop PC

with regular graphics hardware, a keyboard and a mouse. Lastly, from the implementation point of view, a scatterplot is nothing more than a set of simple solid 3D shapes in space. This type of scene is straightforward to implement on most 3D development platforms.

Despite its relative simplicity compared to other visualisation techniques, a 3D scatterplot is a richer and more complex stimulus than the ones commonly used in psychophysical experiments. Consequently, the choice of 3D scatterplot as an experimental stimulus requires a deeper exploration of its characteristic features. The following section will provide an overview of the issues related to rendering a 3D scatterplot on a 2D display.

Rendering an interactive 3D scatterplot

In general, rendering a 3D scene on a 2D medium is a task that is non-trivial, but has been widely researched and practiced by visual artists since the Stone Age, and especially during the Renaissance (Gombrich & Gombrich, 1995). The way to accomplish this is to present sufficient depth cues in the 2D rendering so that the viewer forms the impression of the 3D scene. Not every depth cue can be rendered on every medium, and some depth cues, such as the change of focus distance according to pupil accommodation, cannot be represented at all, at least not with the current technology. A static 2D medium such as a painting or a photograph can incorporate depth cues of occlusion and perspective, as well as weaker cues such as fogging or shift-to-blue with distance (e.g. Goldstein, 2009). It is important to note that perspective as a depth cue relies on objects having a known and recognisable relative size, and is thus of little use in abstract scenes. Animated, but non-interactive 2D media such as cinema add the depth cue of relative motion. Depth cues provided by interactive media such as computer-aided visualisation tools, being recent, are necessarily less researched. Still, it is possible to hypothesise that the visuo-motor correspondence of interaction to stimulus is a strong depth cue in itself. Less common technologies such as stereo displays and head-tracking allow for additional depth cues such as the inter-ocular disparity and a realistic effect of the viewer's self motion on the scene. Nevertheless, with the current state of technology it

is impossible to render all the depth cues perfectly. For example, available stereo rendering invariably creates cue conflicts and often leads to cybersickness (LaViola Jr, 2000).

Usually, in an artificial display, there are contradictory cues to object depth. For example, in an oil painting, perspective hints at depth while pupil accommodation gives evidence that all the depicted objects are in fact situated on the same plane of the painting. It has been repeatedly demonstrated that rendering several coherent depth cues reinforces and facilitates depth perception, while missing or contradictory depth cues result in a situation called cue conflict (Blake & Sekuler, 2005). Research of the effects of cue conflicts shows that they are usually resolved in favour of one set of coherent cues. For example, it has been shown that interocular disparity can override shape cues provided by shading (Cutting & Vishton, 1995). That research, however, does not sufficiently extend to the visualisation techniques commonly used in Visual Analytics (Wickens, et al., 1994). Ware (2000), specifically notes that *“There has been little or no empirical work on the role of depth cues in perceiving structures such as clusters and correlations in 3D.”*

Of the depth cues made available by a 3D scatterplot, perspective, as noted above, is of little use with neither known object sizes nor visually continuous guidelines. Fogging and fading to blue are not strong enough, especially in a shallow depth range. Occlusion is a strong cue as to what object is in front of another, but it does not provide an absolute depth placement information. Moreover, occlusion is dependent on the density of objects on the scene and therefore is not consistently reliable. In fact, a static rendering of a 3D scatterplot does not present strong, consistent and coherent depth cues to enable efficient reconstruction of a 3D scene. This statement is further supported by evidence from common practices: a 3D scatterplot is almost never used in a printed medium (Bigwood & Spore, 2003). Moreover, an influential textbook on multidimensional analysis presents an anecdotal evidence by using only 2D scatterplots for illustrations (Tabachnik & Fidell, 2006).

A dynamic and interactive medium such as a computer display with keyboard and mouse, while still being two-dimensional, provides two additional depth cues that are both strong and consistent: relative motion and visuo-motor correspondence. A cue of relative motion can be easily generated by either moving or rotating the viewpoint around the displayed scene; interactive control of the viewpoint adds the visuo-motor correspondence cue. Several established algorithms exist for both interactive and non-interactive viewpoint control (Ware & Osborne, 1990). These algorithms use UI metaphors of either camera control (“eye in hand”) navigation (“vehicle control”) or manipulation (“scene in hand”). In the most basic example of navigation metaphor, the view rotates around one point imitating the actions of a person rotating their head to take in the surroundings. In a basic example of the manipulation metaphor, the view point instead revolves around the centre of interest, imitating the rotation of an object in an observer’s hands to see it from all angles.

A manipulation algorithm that satisfies the requirements outlined above is a constant-speed rotation of a scatterplot around a constant axis. Interactive algorithms of viewpoint control make use of either standard UI hardware, i.e. keyboard and mouse, or hardware specifically designed for 3D interaction. This multitude of ways to render a 3D scatterplot raises a number of applied questions, including whether the additional depth cues of relative motion and visuo-motor correspondence allow for efficient 3D scene reconstruction; whether interactive scene control improves 3D scene reconstruction; and what effect dedicated 3D interaction hardware has on 3D scene perception. All these are important questions in their own right and will be addressed in the experimental part of thesis.

Experimental methodology

This study assesses human performance in Visual Analytics tasks. The tasks used for assessment are outlier, cluster and trend detection, the most common ones according to the preliminary study (Experiment #0), as described in Chapter 3. The aim of this study is two-fold: validating the use of ‘pop-out’ methodology as a quality criterion for visualisations; and

applying this criterion to the assessment of a specific visualisation technique. A more general and long-term objective is to establish a foundation for a theoretical cognitive account of the processes involved in visualisation comprehension, but achieving that aim exceeds the scope of one thesis.

The 3D aspect of the scatterplot is made visible to the observer using relative motion, by rotating the scene on a screen. The study comprises a series of four experiments that will be denoted in the following sections as '*Experiment #1-#4*'. The methods of these experiments have much in common, so it is parsimonious to describe the general approach here, and address the specifics in the description of each experiment in turn. The detailed rationale and method for each experiment are presented in Chapters 6 to 9 respectively, and summarised in Chapter 10.

Options of interactive rotation control are explored and contrasted in the experiments: automatic vs. interactive, constrained vs. unconstrained. The main analysis, following the 'pop-out' effect paradigm, comprises relating response time to the number of data points. Additional analyses consider accuracy, specific interactivity patterns, and relate the performance of the participants in the experiment to their general profile such as age, gender, basic cognitive abilities etc.

Graphically a 3D scatterplot is a set of spheres positioned in space. Formally, it is a rendering of a data set comprising five variables, four continuous and one discrete. These variables are visualised as a position in space (X, Y and Z displacements), size and colour of a sphere respectively. The data sets used in all the experiments of this study are artificially generated to meet the requirements of each experiment design. Procedures for generating data sets are described in the Method sections of the relevant experiment chapters.

There are many minor considerations involved in various aspects of experiment design. One of these is the possibility of attention lapses (Norman & Shallice, 2000) in performing repetitive tasks over a period of time. One of the ways to maintain participants' attention and

involvement on the task is by providing performance feedback. Auditory stimuli easily elicit emotional responses (LeDoux, 1996). Also, since the main experimental task is purely visio-motor, an auditory feedback should not interfere with the task. Thus, two short sound samples, one high and harmonious, another low and discordant, were prepared and used as feedback throughout all the experiments. They were sounded after each trial, signifying successful or unsuccessful performance.

The results of each of the four experiments are analysed in three stages. The first stage is data screening, including detection and removal of outlier data points, either on a per-trial or a per-participant level, and remapping, aggregating and recalculating the data where necessary. The second stage is the main analyses, i.e. the analyses which answer the main question of the experiment. In most experiments the main analysis is for the ‘pop-out’ effect: a linear regression relating response time to the number of data points. The third and last stage is the secondary and post-hoc analyses. Secondary analyses differ from experiment to experiment and are detailed in the relevant sections of the chapters describing each experiment.

A minor improvement is introduced in the main analysis in all experiments with respect to the evaluation of the effect of data set size on response time. Usually, experiments of ‘pop-out’ effect (Hershler & Hochstein, 2005; Hochstein & Ahissar, 2002; Pomerantz, 2006; Pomerantz, et al., 1977; Treisman, 1982; Treisman & Gelade, 1980) analyse response time as recorded. However, several studies have demonstrated that response time of human subjects is not distributed normally (Ratcliff, 1979; Zaitsev & Skorik, 2002). That violates the normalcy assumption of most standard statistical analyses (Langdridge & Hagger-Johnson, 2009), specifically of the linear regression analysis that is the central tool of ‘pop-out’ effect paradigm.

Characteristic features of the human response time distribution are positive-only values, high negative skew, cut-off to the left and a long trail of outliers to the right of the mean (Zaitsev & Skorik, 2002). Several theoretical distributions that fit that distribution have been proposed.

For example, (Zaitsev & Skorik, 2002) propose a distribution function (Equation 1) that provides a very good fit. Linking from it to normal distribution is, however, relatively complicated since it includes several free parameters that need to be fitted. A log-normal distribution (Limpert, Stahel, & Abbt, 2001) exhibits the same characteristic features, and has the benefit of a very simple link function with no free variables (Equation 2).

$$f = \frac{\lambda}{a} \exp \left[-\exp \left(\frac{b-x}{a} \right) + \left(\frac{b-x}{a} \right) \right]$$

Equation 1 - distribution of human response times according to (Zaitsev & Skorik, 2002)

$$f = \ln(x)$$

Equation 2 – lognormal link function (Limpert, et al., 2001)

In this research, the distribution of reaction times is tested for normality in each experiment starting from Experiment #2, and if it fits lognormal distribution better than the normal distribution, a natural logarithm link function is applied to the data prior to subsequent analyses. This is expected to improve the statistical validity of the tests such as ANOVA and linear regression that assume a normal distribution of the dependent variable. Likewise, and for the same reasons, a logit link function (Cramer, 2003) is used for the analysis of error rates where applicable.

Ethics considerations

The research in this thesis includes running experiments with human participants. Consequently, ethical issues have to be taken into consideration. The participants in all experiments are adults, with no specific requirements or disabilities. There is no covert manipulation of participants in any stage of any experiment. The only foreseeable hazard is due to prolonged exposure to computer monitor. To address that, in experiments which lasted for more than 10 minutes per participant (Experiment #2-4), regular breaks are scheduled during which participants are specifically instructed to rest their eyes. In all the

experiments, participants are fully briefed and debriefed, and sign an informed consent form. The data is anonymised at the collection stage. In accordance with the University regulations, ethical approval was granted by the ethics committee of the School of Social and Health Sciences (see the scanned document on the enclosed CD).

Technical solution

Hardware

The choice of hardware platform for Experiment #1 to #3 was driven by how widespread it is, and by its availability. The underlying assumption was that results received on a widely-used hardware platform would be easy to generalise and integrate into common visualisation practices. It was also expected that participants' familiarity with the common hardware as well as the ergonomic considerations involved in its design would minimise the effects of unfamiliar and uncomfortable interfaces on performance. With the exception of Experiment #4 which specifically examined the effect of a novel 3D interactive device.

Experiments #1-#3 were run on an office desktop PC by Hewlett-Packard, with 2.4GHz processor and 2 Gb RAM. Graphics was rendered with NVidia GeForce 8800 GTS card. The experiments were run in normal daylight, using normal office hardware, with the default colour and 3D settings. Experiment #4 was run in the UAD Human Interactive Virtual Environment suite (HIVE). At the time of the study, the HIVE comprised two 2mx3m screens, passive stereo back-projection system, 5.1 surround sound system and four wired InterSense⁶ 6 degrees-of-freedom position/orientation trackers (head, two hands and a 3D pointing device – a wand). InterSense system reported position (X, Y and Z) and orientation (yaw, pitch and roll) for each tracker with sub-millimetre and microsecond precision. The wand also had five general-purpose buttons: a trigger button, four coloured response buttons (red, green, blue and yellow) and a mini-joystick button that registered variable XY

⁶ <http://www.intersense.com/>

declination as well as the presses. The hardware was connected to a single Dell PC with an NVidia QuadroFX 4600 graphics card.

Software

Software development closely followed the life cycle of the experiments. For every experiment, a dedicated application was written that presented stimuli and logged responses. At the same time, synthetic datasets to be used as stimuli for the experiment were generated and tested. After the experiment was run on participants, the raw data of experimental results were prepared for statistical analyses: reformatted, cleared of outliers, and sometimes re-scaled or recalculated. Lastly, the analyses were run on the prepared data. All the offline processing, i.e., dataset generation, data preparation and analysis, also required dedicated software programming, although it was a small task in comparison with the main application that ran the experiment.

The choice of Java programming language⁷ as the programming platform for experiments was driven by several considerations. First, Java is widely adopted in industry and education, unlike specialised visualisation packages such as VTK⁸. Second, Java, being a 4th generation language, allows rapid prototyping, which makes it a better choice for the development cycle of experimental software than the more computationally efficient platforms such as C or C++. Lastly, Java's rich and generic implementation of both 3D and interactivity is well-suited for the specific experiments planned. Specifically, *java3d* package⁹ is a rendering framework that allows 3D scene rendering for various configurations of screen sizes and positions, stereo rendering and head tracking, in a way that is transparent to the scene-generating software and can be specified externally via a simple configuration file. This package fitted well with experimental requirements.

⁷ <http://www.java.com/>

⁸ <http://www.vtk.org/>

⁹ <http://java.sun.com/javase/technologies/desktop/java3d/>

Experiment #4, undertaken in the HIVE, required writing a device driver interface, in order to connect InterSense tracker device drivers, available as Dynamic Linked Libraries (DLLs), to Java. The driver was based on open-source Java drivers for Polhemus trackers that have similar hardware interface¹⁰, written using JNA (Java Native Applications) technology. Some specific adjustments were made to the open-source drivers to account for the fact that the sensor strips in the HIVE are positioned vertically on the screen itself and not in the default position, i.e. horizontally on the ceiling.

All the established statistical tests, for main, secondary and post-hoc analyses, were done on SPSS using standard statistical routines. Matlab¹¹ programming language and development environment were used for generating synthetic datasets and for preliminary processing of results' data. Matlab was chosen because it is a fast-turnover scripting language enabling tight development cycle. Also, Matlab is geared towards high-performance mathematical operations on large arrays and has a very simple and efficient file input/output system. Matlab scripts that generated synthetic datasets according to experiment design requirements were written for each experiment. The resulting datasets were stored in CSV files (comma-separated values). The CSV files had a row for each data point, detailing its XYZ position, size and (in Experiments #1 and #2) colour. A list of datasets to be used in a single session of training or testing was stored in a protocol file, with the format of protocol file differing between experiments.

The Java applications that were running the four experiments were built on the same object design, based on two main singleton classes: MainForm and Experiment (see Figure 6 for object design). The MainForm class was an extension of JForm class, serving as the main point of entry for the application. The MainForm provided fields for participants' initials and any other per-participant information such as gender, age and field of study (see Figure 7 for a snapshot of Form dialog for Experiment #3 – others were similar). It also provided entry

¹⁰ <http://www.cybergarage.org/vr/>

¹¹ <http://www.mathworks.co.uk/>

points for supplementary computerised tests such as a handedness questionnaire. The other main class, Experiment, was called by MainForm after all the participant details were collected and stored, participant was assigned to a group (in Experiments #2 and #4), and all the output folders were created and ready for logging.

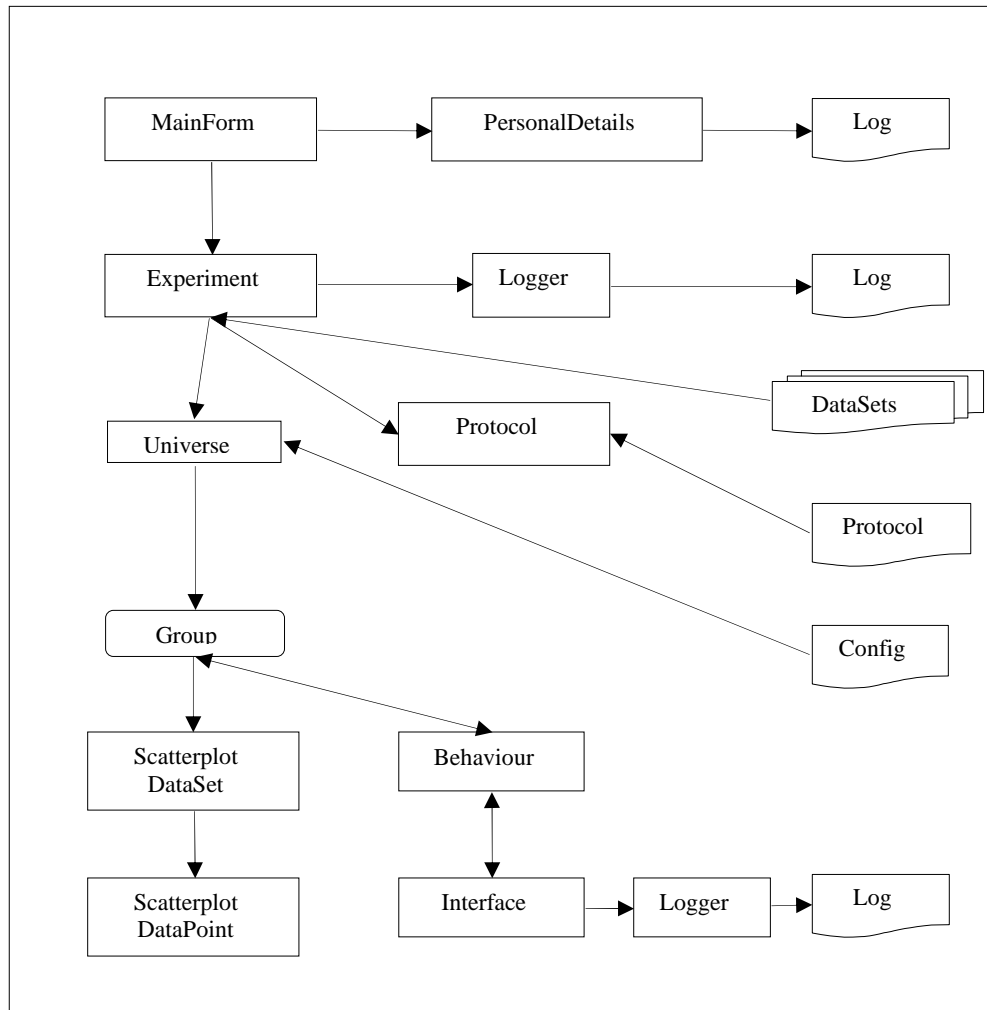


Figure 6 – main application object design. Boxes: classes; rounded boxes: Java library classes; wavy boxes: files.

The Experiment class encapsulated the actual running of the experiment: presenting stimuli and collecting and logging responses. The main parameters passed to Experiment were protocol file name, path name of the folder where all the logs were saved, and the mode to run the trials in: testing or training. In training mode, responses were not logged and the stimuli were presented in a fixed rather than randomised order.

The image shows a Java Swing window titled "Experiment III". Inside the window, there are four rows of input fields and buttons. The first row has an "Initials" label, a text box containing "abc", and a "Run Digit Span" button. The second row has an "Age" label, a spinner box showing "24", and a "Run Training" button. The third row has a "School" label, a dropdown menu showing "SHS", and a "Run Test" button. The fourth row has a "Gender" label, a dropdown menu showing "Male", and no button. At the bottom left of the window is a "Quit" button.

Figure 7 – Experiment #3 MainForm dialog, showing fields for gathering general data about the participant, as well as the controls to run specific tests.

All graphical processing and rendering was managed through a standard java3d class `ConfiguredUniverse`, which managed and rendered the java3d scene-graph. In java3d, the scene-graph is a tree-like data structure describing the scene, with a single root and many branches. Leaves of the scene-graph tree are usually renderable objects, such as polygonal meshes, particles or light sources, but can also be viewpoints, behaviours etc. Between the root and the leaves, the tree is organised using objects of Group family of classes, either the pre-supplied `BranchGroup` or `TransformGroup` classes, or custom-written classes that are derived from either of these. A Group object manages a set of either leaf objects, or other Group objects, thus creating the tree structure. In addition, `BranchGroup` allows adding, removing or modifying its objects in runtime. `TransformGroup`, in addition to leaf and group objects, holds a transform, a data structure that describes position and orientation in 3D space. The position and orientation apply to all the objects that are below the `TransformGroup` in the scenegraph¹². Moreover, the transform is local and relative to the position and orientation of the `TransformGroup` that is higher in the tree. Thus 3D placement of a leaf node is defined by a sequential application of all the transforms between it and the root of the tree.

¹² in computer science, a tree is customarily referred to as having the root on top and the leaves on bottom

The major BranchGroups of the scene-graph in the experiment applications were the scatterplot itself; the light illuminating the scatterplot; and the behaviour controlling the rotation of the viewpoint around the scatterplot. In Experiments #2 and #3, visible interactivity controls were also present as part of the scene. In Experiment #1 no interactivity was used and thus no controls necessary, while in Experiment #4 the controls were implicit in the device, with no need for them being present in the scene. Lighting was set to default, with one source of ambient white light and one source of directional white light from above and behind the camera (45deg in YZ plane). Scatterplot was implemented using custom-written classes `ScatterPlotPoint` and `ScatterPlotPointSet` that extended the standard class `TransformGroup`. `ScatterPlotPoint` encapsulated the application-specific functionality of a single scatterplot point, i.e. a single coloured sphere positioned at given offset from the centre of the scatterplot. `ScatterPlotPointSet` was a transform group including a set of `ScatterPlotPoint` objects, and the rotation of the whole set around an axis. Since `ScatterPlotPoints` were in the sub tree under `ScatterPlotPointSet`, the transform of the `ScatterPlotPointSet` determined the orientation of the whole scatterplot, while the transform of each `ScatterPlotPoint` determined the position of that point relative to all other points.

The transform of the `ScatterPlotPointSet` was controlled by the custom-written `Behaviour` class, an implementation of java3d behaviour interface, which received and processed the relevant interface and timer events. In different experiments the interface events were mouse or 3D wand movements, mouse clicks, keyboard presses or wand button presses. `Behaviour` was invoked before every frame was rendered, and updated the orientation of the scatterplot according to the input. This class also time-stamped and logged every processed interface event. In Experiments #2 and #3, the `Behaviour` class also included control of the appearance of the interaction controls and buttons.

The logging structure implemented resulted in several levels of experiment logs with increasing resolution, organised in a folder tree (see Figure 8). At the top level, experiment log kept track of each participant's details. For each participant, a separate log was kept

detailing the order of trials presented to this participant and the outcomes of each trial. Lastly, for each trial, an interface log tracked every action with millisecond resolution. During the processing of the results, interface logs were processed to yield trial-level variables (see Analysis of Experiment #2 for details) and these were integrated into per-participant log. All per-participant logs were then aggregated into a single table for analysis.

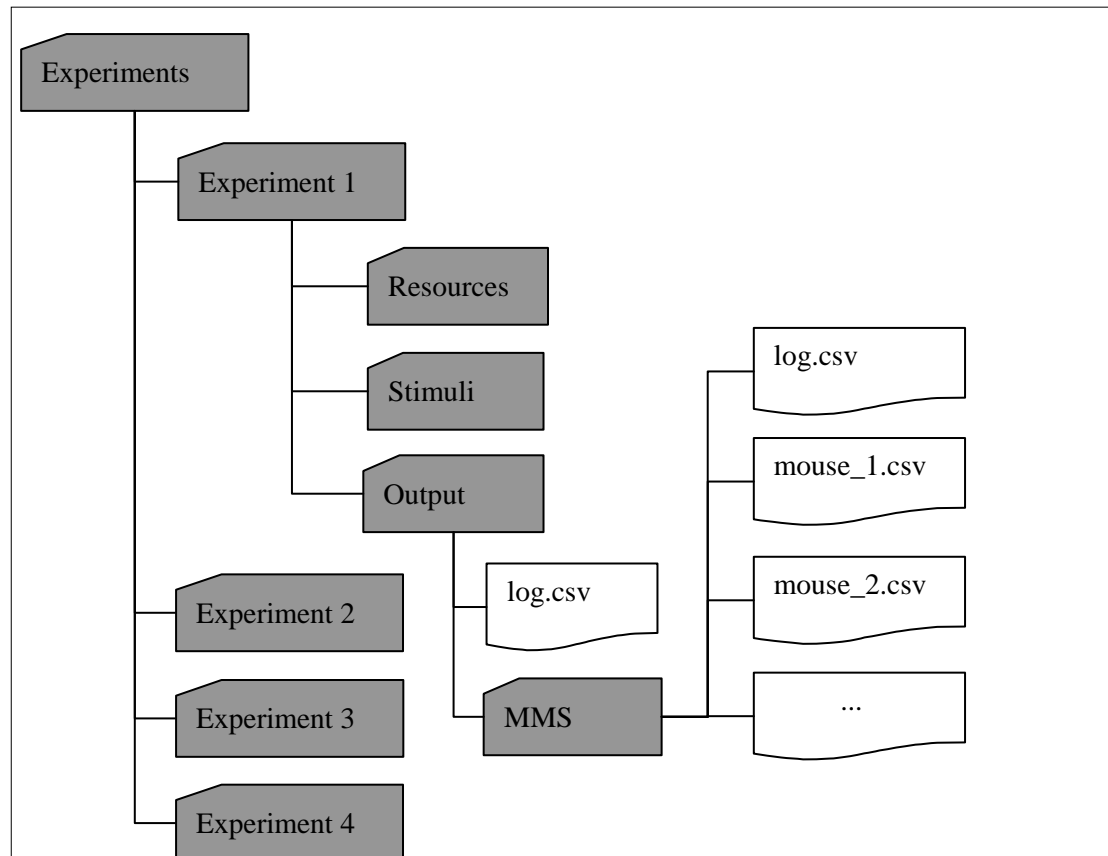


Figure 8 – file structure for experiment resources and logs. Gray boxes: folders; white boxes: files.

The next four chapters apply the methodology outlined in this chapter to provide experimental evidence that forms the central part of the thesis. Each chapter presents one experiment, in the following form: Introduction, Method, Results and Conclusions. Method will mostly present the differences from the overall methodology presented in this chapter, and interim Conclusions will focus on the specific lessons learned that affect consequent experiments. An overall discussion collating outcomes of the experiments will be presented separately in Chapter 10.

Chapter 6. Experiment 1: outlier detection in a 3D scatterplot

Introduction and overview¹³

This experiment demonstrates the feasibility of applying ‘pop-out’ methodology to dynamic stimuli. The ‘pop-out’ methodology consists of measuring reaction response time in an odd-one-out detection task while manipulating the size of the data set. It is hypothesised that the pattern of results reveals the processes of high-level semantic parsing of the scene, i.e. information retrieval. The experimental stimuli required to test the predictions of this thesis are sufficiently different from the stimuli typically employed in ‘pop-out’ experiments (cf. those reviewed in Chapter 2) to require such a proof. There are two main differences: first, the stimuli – 3D scatterplots – are more complex than those used in the experiments that laid the foundations for the Reverse Hierarchy Theory (Hochstein & Ahissar, 2002); second, the stimuli are dynamic, and experimental tasks involve interaction with the stimuli. This could potentially affect the main measured variable, the response time, introducing factors related to scene dynamics and interactivity that are not directly related to target perception. To test the feasibility of ‘pop-out’ methodology in these conditions, a simple experiment was designed and run. It consists of a single task: single outlier detection, with outlier defined by a single depth cue, with straightforward experimental design and minimal interactivity.

Even in a design as simple as that, a complex stimulus such as a 3D scatterplot has many free parameters, such as colour, shape and size of the scatterplot points, 3D distribution of distracter points and the position of the target relative to distracters. The effect all these free parameters have on performance is unclear, and one of the aims of this feasibility study is to establish reasonable value ranges for these. The position of the target relative to the

¹³ The results of this experiment were presented at European Conference on Visual Perception 2008 (see appendices and the attached CD for details)

distracters is the most interesting parameter from the applied point of view: it is important to know how different an outlier has to be before it is detected in a particular visualisation technique. Therefore in this experiment it is varied between two options: either relatively far from the rest of the distracters or relatively close, with the prediction being that the outliers that are further away from the distracters will be easier to detect.

A pilot test has shown that occlusion of the distracter sphere by the target resulted in immediate perception of the target, overriding relative motion depth cues. Therefore in the actual experiment the datasets were pre-recorded and checked for occlusions, and only the datasets without occlusions were used. The pilot study also emphasised the importance of having the touch screen securely mounted and regularly cleaned. Accordingly, during the experiment itself the screen was fixed to a wall, thus making it easier to press and impossible to move, and it was cleaned of fingerprints between participants.

The hypothesis tested in this experiment is that ‘pop-out’ methodology produces results that are consistent with the Reverse Hierarchy Theory when applied to complex, dynamic stimuli. The specific predictions are as follows: first, that the relation between response times for outlier detection and stimulus set size is either constant or linearly increasing, but not decreasing or non-linear; and second, that response times for targets that are closer to distracters are higher than for the targets that are further away. The study uses a within-subject design with three levels of array size and two levels of outlier displacement, measuring one dependent variable: mean response time for correct answers.

Method

Apparatus

The experiment was run in a quiet office space under standard office lighting conditions. Stimuli were presented and responses collected using NEC 17" MultiSync LCD 175VXM+ touch-screen monitor connected to a standard desktop PC running Windows XP. The

touchscreen was the only input/output device throughout the experiment. It was securely mounted to avoid any monitor movement as a result of participants' actions, and it was cleaned regularly to remove any fingerprints. Participants were given an opportunity to adjust their seating and viewing distance to their comfort, thus ensuring that the monitor screen remained well within arm's length at all times.

Procedure

At the beginning of the experiment, participants were given an information leaflet and a consent form (see the enclosed CD). The information leaflet explained that the task in each trial was to touch the sphere that appeared to be positioned closest to the viewer. After signing the consent form, participants undertook a short training exercise to become familiar with the task. The training consisted of 10 pre-generated datasets of increasing complexity, starting with a square grid of spheres of same size and depth with a single sphere positioned in front of the rest. Subsequent datasets gradually introduced random position, random depth and random sphere size, with the last three being of the same design as in the main experiment. The pilot experiment indicated that this training exercise was appropriate to enable ready comprehension of the task.

The experiment comprised 36 trials, with the same stimuli being presented to all participants. The order of presentation was in four blocks providing an even distribution of trials with different set sizes and target depths, with half of the participants receiving the protocol in reverse order, to reduce order effects. Each trial was preceded by a fixation screen, comprising a single sphere rendered in the centre of the screen. The trial started when the participant touched the sphere. This gave participants an opportunity to pace the experiment as well as ensuring consistent positioning of the index finger at the beginning of the trial. After the fixation sphere was pressed, the stimulus was presented until the user touched one of the spheres on the screen, for 5 seconds at the longest. After either response or time-out, the target sphere blinked for 1.5 seconds, providing feedback to the participant.

The duration of the experiment for each participant, including explanations and training, was between five and eight minutes.

Stimuli

Experiment design included six stimulus types: three array size conditions 25, 35 or 55 points per data set, and two target depth conditions: target at either 3 or 5 standard deviations (SD) away from the mean. Six datasets were pre-generated for each of the six conditions, resulting in 36 trials per participant overall. The position of each sphere in the XY plane was randomly drawn from a uniform distribution, covering the central 15x15 cm area of the screen. That resulted in densities of 0.11 items/cm², 0.16 items/cm² and 0.24 items/cm² for stimuli set sizes of 25, 35 and 55 spheres respectively. The depth (Z-axis position) of the spheres was randomly drawn from a normal distribution with mean 0 and SD equivalent¹⁴ to 6 mm, cut off at 2 SD around the mean. Sphere diameters were randomly sampled from a uniform distribution of 7-10mm.

The depth of a single sphere, the outlier target, was changed to position it ‘in front’ of the rest, at either 3 or 5 SD from the mean, which is equivalent to 1.8 cm or 3 cm if, as depicted Figure 9D, the scene would be viewed in 90° rotation. The colour of each sphere was randomly red, green, blue or yellow. After the stimulus arrays were generated, they were tested to ensure that at no time during rotation did any sphere occlude any other, and also that the target sphere was not an extreme value in size, colour or planar distance from the centre, i.e. that it is neither the largest nor the smallest, that its colour is not unique in the scene, and that it does not ‘stick out’ to the sides of the scene from any of the available rotation angles.

¹⁴ ‘Equivalent’ meaning that if the scene was a real 3D stimulus, and not a 2D rendering, then that number would be the distance in the depth dimension.

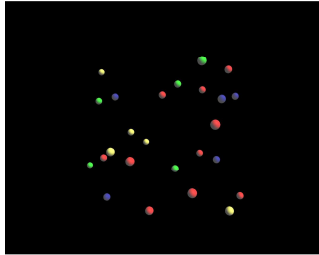
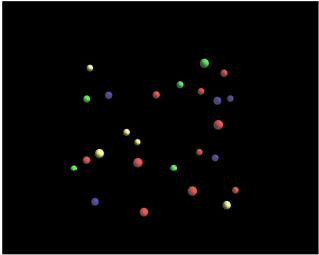
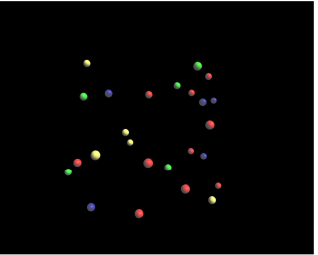
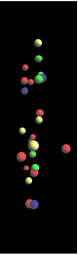
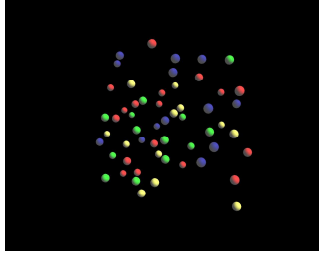
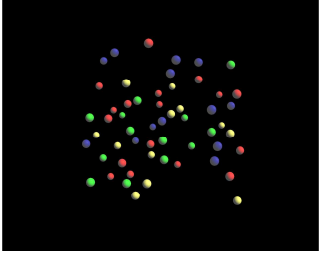
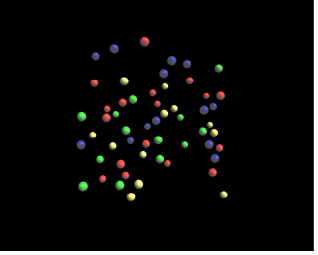
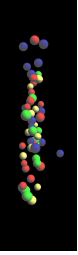
	A	B	C	D
N	Maximum left rotation	Frontal view	Maximum right rotation	Side view
25				
55				

Figure 9 – stimuli used in Experiment #1. Top row: dataset with 25 points; bottom row: dataset with 55 points. Columns A, B & C show different stages of rotation: -22.5° , 0° and 22.5° respectively. Column D presents a side view, rotated 90° around vertical axis, which clearly shows the outlying target. This side view was never shown to the participants.

The stimuli were rendered as two-dimensional animations, with the whole scene rotating around the vertical axis (Figure 9A, B and C). The rotation was at a constant speed of $15^\circ/\text{sec}$, alternating left and right, from -22.5° to 22.5° , i.e. covering an aperture of 45° in 3 seconds. Rendered sphere sizes varied due to rotation (i.e. grew when rotation brought them closer to the front). Because of this, and the fact that each participant's head position was only constrained by the implicit requirement of having the touch screen within comfortable reach, the visual angle subtended by each item was not constant, varying between 0.5 and 1.5 degrees of visual angle. The whole scene was rendered on a black background using the OpenGL standard Blinn-Phong lighting model (Blinn, 1977) with a single white light at viewer position. The diffuse colour component of the shading calculation was the system default red, green blue or yellow, according to the sphere colour; the ambient colour component was 50% grey and the specular colour component was system-default white.

Participants

Participants were selected by convenience sampling from the Environmental Sciences group of the University of Abertay Dundee. This was done to increase sample validity, since the majority of the participants use methods and tools of Visual Analytics in their everyday activities. All participants were fully briefed and debriefed, and signed an informed consent form before commencing the experiment. Twenty five participants (10f/15m) within age range 22-48, median age 40, with normal or corrected to normal vision, took part in the experiment over a data collection period of two days.

Results

Experiment results are summarised in Figure 10. The data were analysed with two univariate general linear model (GLM) routines, separately for correct target identification (binary yes/no) and response time (milliseconds). Response time was analysed for correct answers only; response times for incorrect answers or response times longer than 5 seconds were considered missing values. Outlier depth displacement (either 3 SD or 5 SD), and stimulus array size (25, 35, or 55) were the fixed factors in the model; participant ID entered the model as a random factor.

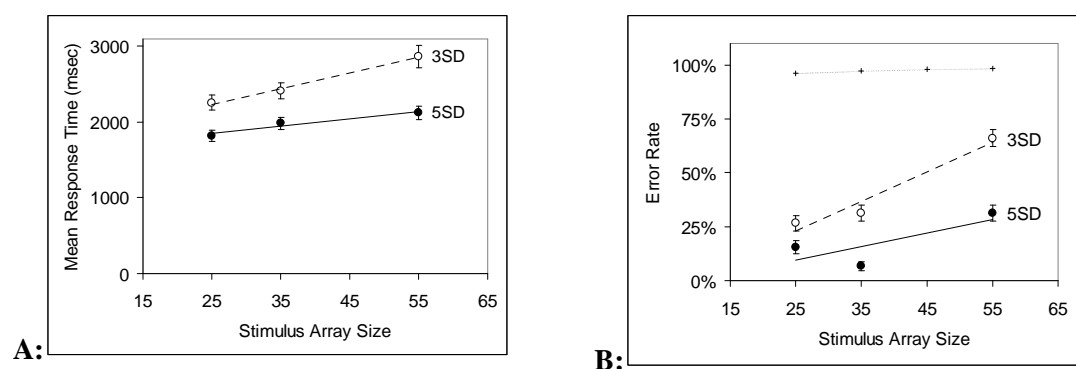


Figure 10 – results of Experiment #1. A: average response times for correct answers; B: error rate. Error bars indicate ± 1 standard error of mean; solid lines and filled circles: target at 5 SD from the mean; dashed lines and empty circles: target at 3 SD from the mean. Greyed dashed line and pluses in chart B is the theoretical chance performance level. Lines are at linear fit to data. N=25.

A GLM analysis for response time showed significant main effects for all the independent variables: stimulus array size $F(2,48)=14.757$, $p<0.001$; outlier depth $F(1,48)=46.047$, $p<0.001$ and between-subject variability: $F(24,48)=9.018$, $p<0.001$. Specifically, decreasing outlier depth from 5 SD to 3 SD increased the response time, on average, from 2.16 to 2.44 seconds; increasing stimulus array size from 25 to 55 increased average response time from 2.02 to 2.37 seconds. No interaction effects were found to be significant.

A GLM analysis for the ratio of correct answers showed a similar pattern: significant main effects of stimulus array size $F(2,48)=70.499$, $p<0.001$; outlier depth $F(1,48)=99.360$, $p<0.001$ and between-subject variability: $F(24,48)=7.292$, $p<0.001$. In addition, the interaction of outlier depth and stimulus array size was found to be significant: $F(2,48)=9.978$, $p<0.001$. Increasing outlier depth was reflected in an increase in percentage of correct answers from 59% to 82%. When the outlier was positioned at 3 SD, increasing stimulus array size from 25 to 55 led to a decrease of accuracy from 73% to 34%; with the outlier at 5 SD, the decrease in accuracy was reduced from 85% to 69%.

A follow up analysis, including parameters of the stimulus that were not part of planned analyses, revealed a more complex picture. A simple regression showed that target size and the distance of the target from the centre in the XY plane have a small but statistically significant association with response time: $\text{adj}R^2=0.02$, $p<0.001$ and $\text{adj}R^2=0.015$, $p=0.001$ respectively. Larger targets, as well as targets closer to the centre of the screen, were identified faster than those smaller or further away. The inclusion of target colour, the only other variable of the target stimulus, in the model did not show any effects of colour on either speed or accuracy of performance.

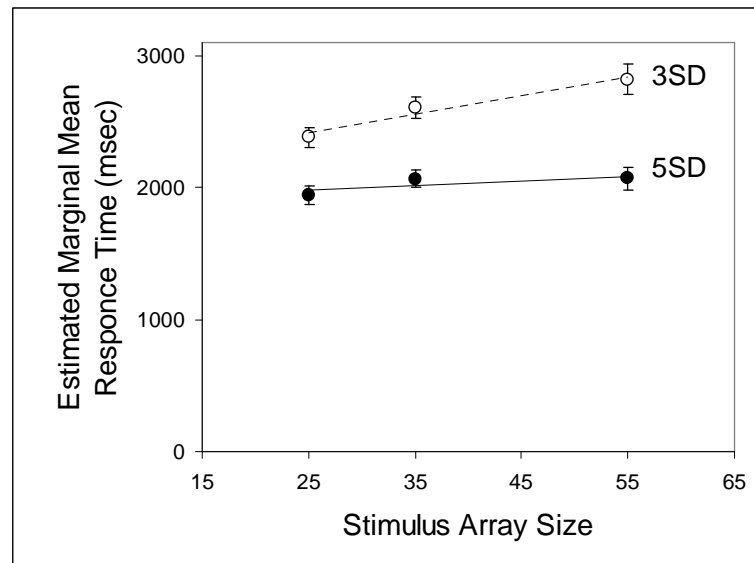


Figure 11 – estimated marginal means of response times for correct answers after correcting for target size and distance from centre. Error bars indicate ± 1 standard error of mean; solid lines and filled circles: target at 5 SD from the mean; dashed lines and empty circles: target at 3 SD from the mean. Lines are at linear fit to data. $N=25$.

To control for the effect of target size and distance from centre, these variables have been included in the main GLM as covariates. A plot of estimated marginal means, corrected for these covariates, is shown in Figure 11. This augmented analysis again showed significant main effects for all the variables included: stimulus array size: $F(2,485)=5.992$, $p=0.004$; target depth displacement: $F(1,486)=53.948$, $p<0.001$; target size: $F(1,486)=21.989$, $p<0.001$; and the distance of the target from centre: $F(1,486)=11.270$, $p=0.001$. Between-subject variability was also significant: $F(24,463)=9.642$, $p<0.001$. The estimate, corrected for both covariates, is that with the 5 SD outlier, increasing set size from 25 to 55 increases response time from 1.95 to 2.07 seconds (a 6% change). With the 3 SD outlier, the same increase in set size increases response time from 2.38 to 2.82 seconds (an 18% change). A separate GLM analysis of response time for 3 SD and 5 SD outliers, corrected for target size and distance from centre, showed significant effects of stimulus array size only for the 3 SD outliers $F(2,485)=5.039$, $p=0.009$. For 5 SD outliers, the effects of stimulus array size were not significant: $F(2,485)=1.311$, $p=0.276$.

Lastly, to account for the possibility of the incorrect answers being caused by participants missing the target sphere, an analysis of incorrect answers was done, showing that the

distance, on the screen, of an incorrect response from the target was, on the average, 3.42cm, with standard deviation of 1.4cm. This makes it unlikely that the wrong answers were caused by participants missing the target sphere on the touch screen and hitting a neighbouring sphere by mistake.

Discussion

Overall, the results of Experiment #1 were in line with the predictions formulated at the design stage. As expected, targets that were further away from the distracters in depth (5 SD condition) were detected more easily both in terms of speed and accuracy. Response time for correct answers increased linearly with the number of distracters, and even the unexpectedly strong influence of target size and distance from the origin did not obscure the main effect of the dataset size. This permits interpretation of the results in terms of the ‘pop-out’ paradigm and the Reverse Hierarchy Theory. In these terms, the results indicate that depth-defined targets in a 3D scatterplot are detected through serial visual search, and not through bottom-up processes of immediate perception. The applied implications are that this particular visualisation technique is not efficient for outlier detection. The applied, methodological and theoretical implications of Experiment #1 are discussed in full detail in Chapter 10, in conjunction with the rest of the empirical material from this thesis and in the context of the existing literature.

However, serial visual search is not the only possible explanation of the pattern in the results from Experiment #1; at least two other interpretations are possible. One is based on the fact that the underlying assumption of the visual search interpretations is that the participants complete the task of outlier detection by perceptually parsing the stimulus into a 3D scene; a scene that consists of a set of spheres lying roughly on a vertical plane, with the target sphere ‘sticking out’ in front of it. However, since the ‘reference’ plane in all the stimuli is vertical, the target is also an outlier in another parameter that is, possibly, more easily detected: it moves with the highest local speed across the screen compared to its neighbours.

It is, therefore, possible to identify the target simply by comparing local relative motions of spheres. Since the design of Experiment #1 did not make possible the dissociation between local relative motion and global 3D scene reconstruction, the validity of an interpretation of the results based on 3D visual search is an open question, at least until subsequent experiments provide evidence that can dissociate between the interpretations.

Another competing interpretation is based on Fitts law (Fitts, 1954). A prediction derived from this law would be that motor response time is affected by target size, target distance, and, also, by the number of distracters. The first two predictions were indeed supported by a *post-hoc* analysis. This experiment alone cannot determine whether stimulus array size affects target identification time or just motor response time, since the measured variable is the overall time, including the time it takes to detect the outlier as well as the time it takes to physically move the finger to the outlier location on the screen and press it. Again, subsequent experiments must be designed to dissociate between these interpretations.

Conclusions and implications

On a methodological level, the experiment shows that the ‘pop-out’ methodology is a feasible tool for measuring performance in visualisation tasks, providing interpretable results for complex animated stimuli. Thus the theory of perceptual organisation appears to be a useful framework of defining and assessing visualisation usability. Most importantly, the results and analysis of Experiment #1 highlighted several issues that influence the design of subsequent experiments. The first and main issue is the need to dissociate between three competing interpretations of what specific process is affected by increasing array size: is it the 3D scene reconstruction; the local 3D motion; or the motor response planning. Second, since target colour does not seem to affect performance, the same system default colour scheme can be used in the next experiments. Finally, since the effect was observable at both 3 SD and 5 SD conditions, subsequent experiments can concentrate on the more practically relevant case only, with an outlier at 3 SD from the mean.

Chapter 7. Experiment 2: the role of interactive view control in outlier detection

Introduction¹⁵

Experiment #2 continues the line of enquiry established in Experiment #1, focusing on the effects of interactivity on outlier detection in a 3D scatterplot. During Experiment #1 several participants commented that the constant rotation of the scene was not efficient, and that they would prefer to be able to change the view themselves. Most 3D visualisation tools currently in use give users some interactive control of the view, enabling adjustment of the position and orientation of the viewpoint relative to the scene and the viewing angle¹⁶. In Experiment #2 these two modes of viewpoint control: passive, with the scene constantly rotating in front of the viewer, and active, with camera control in viewers' hands, are directly contrasted. The experimental question is, again, on two levels: methodological, whether 'pop-out' methodology can be applied to interactive tasks; and applied, what effect does interactivity have on task performance.

Experiment #2 addresses the methodological shortcomings of Experiment #1. While Experiment #1 demonstrated an effect that is consistent with visual search pattern, it did not produce any evidence of a 'pop-out' effect in a complex visualisation task. An indication that 'pop-out' can be observed in some variation of 3D scatterplot outlier detection task would be an important contribution to the methodological level of enquiry. One of the best-documented 'pop-out' effects is for objects that differ from the rest in colour (Pomerantz, 2006; Treisman, 1982). Thus, an outlier detection task in Experiment #2 is composite, with an outlier defined by position in most of the trials, and by colour in some.

¹⁵ The results of this experiment were presented at the 13th International Conference on Information Visualisation 2009, and at Psychology Postgraduate Action Group conference 2009 (see appendices and the attached CD for details)

¹⁶ In computer graphics, the convention is to refer to these as the camera parameters.

Another shortcoming of Experiment #1 design was that the task of depth outlier detection in a vertical plane did not allow the distinction between local relative motion and global 3D scene reconstruction. The distracter points were positioned along a vertical plane and rotated around a vertical axis. The target was positioned far from the reference plane, and so also from the rotation axis, resulting in a target having significantly different planar speed across the screen from the distracters. To address this, in Experiment #2 the reference plane is dissociated from the axis of rotation. The axis of rotation remains vertical, while the reference plane is tilted 45° away from the viewer (see Figure 12). In half of the trials the tilt is with the top part away from the viewer ('floor' slant), in the other half the top is closer to the viewer ('ceiling' slant). This creates subsets of points, either in the top or bottom part of the data set, with the same depth as the target. The local relative motion of these points was also equal to or greater than that of the target, making outlier identification from local motion impossible.

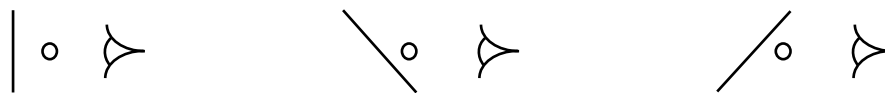


Figure 12 – schematic position of the observer's view point (symbol of an eye), the stimulus reference plane (line) and the target (circle), in Experiment #1 (left); Experiment #2 'floor' slant (middle) and 'ceiling' slant (right); side view.

The use of touch screen in Experiment #1 presented several problems: the participants' hand, while reaching to respond, was partly obscuring the view; the use of a touch screen limited performance measurement to a single recordable event: the time and position of a finger press. This manner of response recording did not allow differentiation between different stages of task performance, in particular between interaction stage and motor response stage. The change of response device from touch screen to mouse, and specifically recording the time and position of a mouse pointer with high enough temporal resolution, solves all these problems. In addition, using a mouse pointer as an input allows for greater precision than using a finger. Thus the size of each sphere can be decreased, and

consequently a wider range of data set sizes can be tested than in Experiment #1, in line with the range of realistic data set sizes (e.g. Fekete & Plaisant, 2002).

In Experiment #1, in order to improve generalisability of the research results with respect to common Visual Analytics practices, the point of view of participants was not constrained, with the only implicit restriction being the arm length requirement imposed by the use of the touch screen. Since that restriction is removed in Experiment #2, and in order to limit within-and between- participant variability in subtended angle of view of the stimulus, a chin-rest is used in Experiment #2.

Finally, Experiment #2 seeks to address in several ways the problem of between-subject variability that was evident in Experiment #1 results. First, a chin rest is used to ensure that the viewing conditions were the same for all participants. Also, the general profile of the participants is assessed for later comparison with performance indicators. This profile focuses on variables that are easy to measure and can be reasonably assumed to influence performance: age group, sex, occupation, left- or right-handedness and memory span. The latter is assessed by a computerised adaptation of Wechsler digit span test (Wechsler, 1997) that is considered by many researchers a reliable, if coarse, measure of general cognitive abilities (Fry & Hale, 1996). While incomparable with the Wechsler version, the computerised version can be assumed to have the required construct validity, in particular concurrent validity. Specifically, between-participant differences in memory span can be expected to be reflected in differences in performance in the computerised version of the test to the same extent as the traditional verbal one.

To summarise, this experiment tests several hypotheses, listed below. The specific predictions based on these hypotheses are presented at the end of Methods section, after the detailed description of experiment design.

1. ‘Pop-out’ methodology produces results that are interpretable within the proposed theoretical account when used in a task that requires interaction as part of scene perception stage.
2. Interactivity impacts task performance pattern.
3. ‘Pop-out’ effect is observable in some variant of a scatterplot outlier detection task.
4. 3D scatterplot outlier detection task is based on 3D scene reconstruction rather than on local motion comparison.
5. In the case of serial visual search, number of elements affects the time required to identify the outlier, but not the time required to make a motor response.

The experiment design is built on the following two manipulations: an outlier that is defined by either colour or position; and a viewpoint control that is either passive or active. If the outlier is defined by colour, all the points in a scatterplot are one colour and the outlier is the opposite colour in the red-green blue-yellow colour space (Blake & Sekuler, 2005). If the outlier is defined by position, all the points in the scatterplot are positioned roughly along a plane tilted away from the viewer, and one point is in front of this plane (in front only so as not to obscure the target). In passive viewpoint control mode, participants start and stop a constant rotation of the scatterplot with the same rotation parameters as in Experiment #1. In active viewpoint control mode, participants can directly control the view, within the same limits as in the passive mode. This results in a within-subject design with two levels of outlier type, two levels of interactivity, five levels of data set size (to test for possible non-linear effects), and two levels of reference plane slant.

Method

Apparatus

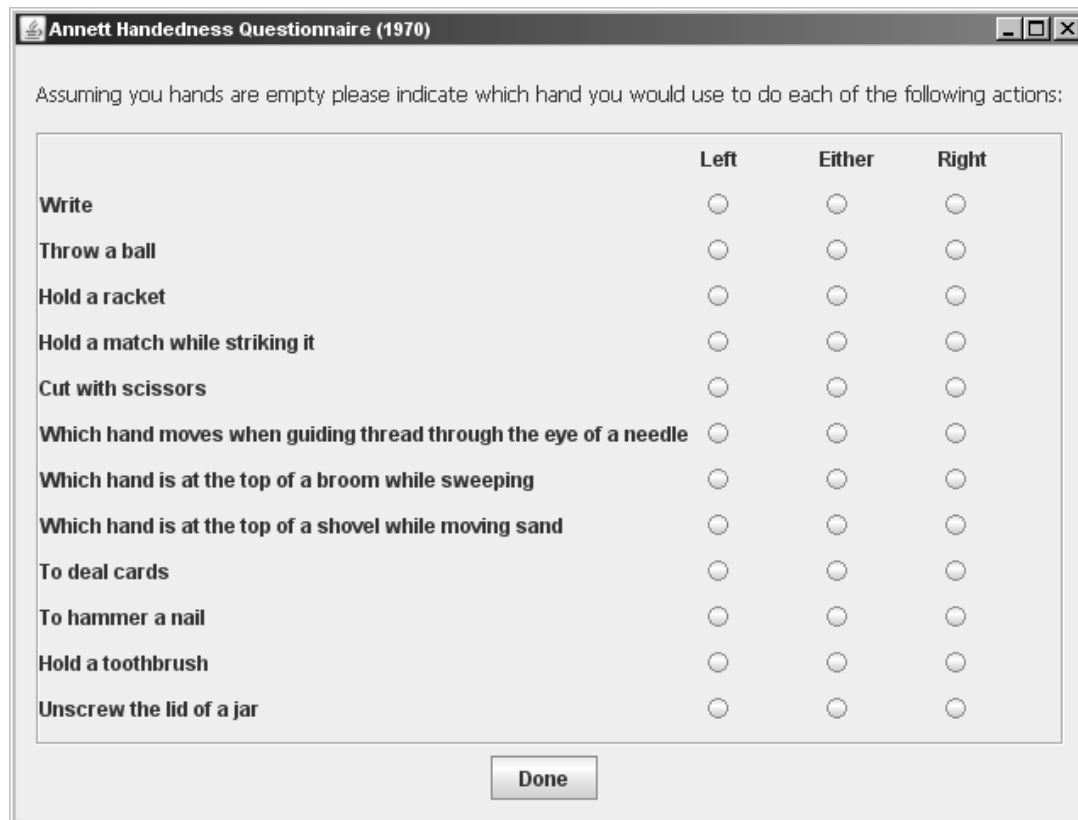
The experiment was run in private office space. Stimuli were presented on a 17” LCD monitor and responses collected using Logitech USB mouse and keyboard. During the main task of the experiment, every mouse event, i.e. mouse movement, button press or button

release, was timed and logged. A chinrest was used to ensure that participants' eyes were positioned 60cm in front of the centre of the screen. Participants were given an opportunity to adjust the height of the chair to their comfort.

Procedure

At the beginning of the experiment, participants were given the information leaflet and a consent form (see the enclosed CD). The information leaflet explained that the task in each trial was to click on the sphere that differed from the rest in either colour or position. After signing the consent form, participants filled in the details in the main dialog form: initials, age range (0-5, 6-10, 11-15, 16-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81+), school (SHS, CCT, CS, DBS or Other) and gender (Male or Female). At this stage they were also assigned to one of four groups in order of arrival.

After filling in the details, the participants filled in a computerised version of a handedness questionnaire (see Figure 13) adapted from Annett (1970), and did a computerised digit span test. In the digit span test, 5mm-high digits were presented in the centre of the screen, black on gray, one after another. Every digit was shown for 300msec followed by 700msec blank. After the end of the sequence, participants were prompted to enter the sequence from memory and the entry form field was enabled. As in the original WAIS version, the length of a sequence started with two digits and increased by one after every two trials, and the test terminated after two wrong answers in a row.



Assuming your hands are empty please indicate which hand you would use to do each of the following actions:

	Left	Either	Right
Write	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Throw a ball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hold a racket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hold a match while striking it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cut with scissors	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which hand moves when guiding thread through the eye of a needle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which hand is at the top of a broom while sweeping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Which hand is at the top of a shovel while moving sand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To deal cards	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To hammer a nail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hold a toothbrush	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unscrew the lid of a jar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Done

Figure 13 – handedness questionnaire form used in Experiment #2.

Before beginning the main experiment task, participants undertook a short training exercise. The training consisted of five sets of trials of increasing complexity, using three sets of stimuli. The first stimuli set comprised four trials where the outliers were defined only by colour, and the number of elements in a data set increased from 35 to 100, 200 and 343. The second training stimuli set comprised 10 trials, with outliers defined only by depth, and difficulty increasing in the same manner as in Experiment #1. The third training stimuli set was of the same complexity as in the actual experiment, comprising 7 trials: two with colour outlier and five with position outlier. Training set 1 used the first stimuli set; training sets 2 and 3 used second stimuli sets, with set 2 in passive and set 3 in active viewpoint control mode; while sets 4 and 5 used the third stimuli sets, likewise running in passive and active control modes respectively.

The experiment comprised 128 trials organised in four sets of 32 trials each. In every set, there were 12 trials with colour outlier and 20 with position outlier. The trials within each

set were presented in random order. The order of set presentation was balanced between participants according to group allocation. Two of the sets were run in active mode, and two in passive. Overall, 64 data sets were pre-generated. Every data set was presented twice to each participant, once in active and once in passive mode. Participants were encouraged to take a break and rest their eyes between sets. The duration of the experiment for each participant, including explanations and training, was approximately 40 minutes.

The user interface in both interaction modes was made as visually uniform as possible, in order to reduce unwanted effects of interface on performance. The interface comprised a gray bar and a square button on it (see Figure 15). The bar was 1378x35pixels, positioned at the bottom centre of the screen, and covered 80% of the width of the screen. The size of the button was 35x35 pixels. The button was showing one of three designs, in either embossed or sunk shape¹⁷ (see Figure 14 for close-up or Figure 15 for the full-screen view of the slider bar in the context of the stimulus). Between trials and in passive rotation mode, the button was positioned in the middle of the bar. In active rotation mode, the button could be dragged along the bar.







	Next Trial	Passive Rotation Control	Active Rotation Control
Released			
Pressed			

Figure 14 – Experiment #2 interface buttons.

Each trial started with no stimulus present, and the button positioned in the middle of the bar and set to ‘Next Trial’ icon. When the participant clicked on the button, the image of the scatterplot appeared in the centre of the screen and the interaction stage began. This gave participants an opportunity to pace the experiment and ensured consistent positioning of the

¹⁷ The author wishes to thank Mr. Robin Sloan for his help with the design of the icons.

mouse pointer at the beginning of the trial. In passive rotation mode, a mouse click on the button toggled the rotation of the scatterplot on and off. The speed, range and axis of the rotation were same as in Experiment #1. In active rotation mode, pressing and dragging the button along the bar rotated the scatterplot; rotation continued until the participant released the button, even if the mouse pointer moved away from the button. This last feature was added after a pilot run showed that the need to keep the mouse pointer on the bar is an attention-demanding task. The position of the button mapped directly to scatterplot rotation angle. The axis and range of rotation were the same as in the passive mode, in order to present the same visual information to participant in both interaction modes.

The interaction stage continued until the participant clicked on one of the data points, however clicks made in the first 250 msec of a trial were not recorded, as they were considered to originate from a preceding trial. In passive rotation mode, the response option was turned off while the scatterplot rotation was on, in order to eliminate mistakes caused by clicking on a rotating scatterplot,. After the response, auditory feedback sounded, the scatterplot disappeared, the button position reset to the centre of the bar and the button icon changed to 'Next Trial'.

Stimuli

64 data sets were pre-generated for this experiment: 24 with colour outlier and 40 with position outlier. Data sets with position outliers were generated in five sizes: 27, 64, 125, 216 and 343 points per set, these being the cubic power of a natural sequence: 3^3 , 4^3 , 5^3 , 6^3 and 7^3 . For colour outliers, in order to reduce the overall number of trials, only three set sizes were used: 27, 125 and 343. The position of each sphere in the XY plane was randomly drawn from a uniform distribution. In contrast to Experiment #1 where the area was kept constant and the density varied, in Experiment #2 the density was kept constant at one point/cm², and thus the area covered by the scatterplot ranged from 5.2 x 5.2 for 27 points to 19x19 cm for 343 points. The size of points was randomly chosen from uniform

distribution, ranging between 1mm and 4mm if seen at zero depth. The target was always 2.5mm in size. The densities in this experiment were much higher and the point sizes much smaller than in Experiment #1. Since the response was made with a mouse rather than a touchscreen, reducing point size did not seem to present a problem.

Point colours and illumination conditions were identical to those used in Experiment #1. In data sets with position outlier, the colour of each point was randomly chosen from the red, green, blue or yellow. In data sets with a colour outlier, in order to maximise the colour contrast, the colour of target was opposite to the colour of the distracters on CIE hue circle (Wyszecki & Stiles, 1967). Thus, red targets were always presented within green outliers only, blue targets – with yellow outliers and vice versa.

In data sets with a colour outlier, all the points were on a vertical plane, with depth set to zero. In data sets with a position outlier, the position of points on Z plane was first drawn from a normal distribution with mean at zero and SD of 1cm, cut off at 2 SD from the mean; the depth of the target point was set at 3cm from zero plane, towards the viewer. Then the whole data set was rotated 45° around X axis, in one of two directions: either slanting away (the ‘floor’ slant), or towards the viewer (the ‘ceiling’ slant). The resultant stimulus arrays are shown in Figure 15 and Figure 16.

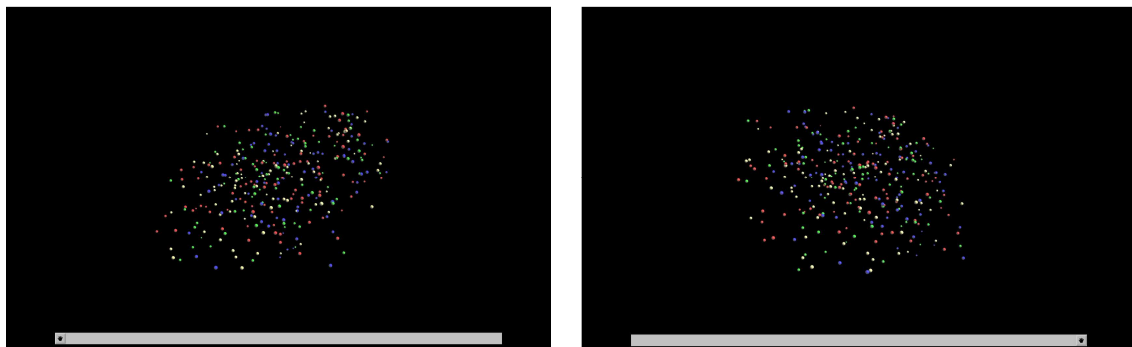


Figure 15 – stimuli used in Experiment #2; a dataset with 343 points with a position outlier, in ‘floor’ slant, in a manual rotation mode, showing left and right rotation limits.

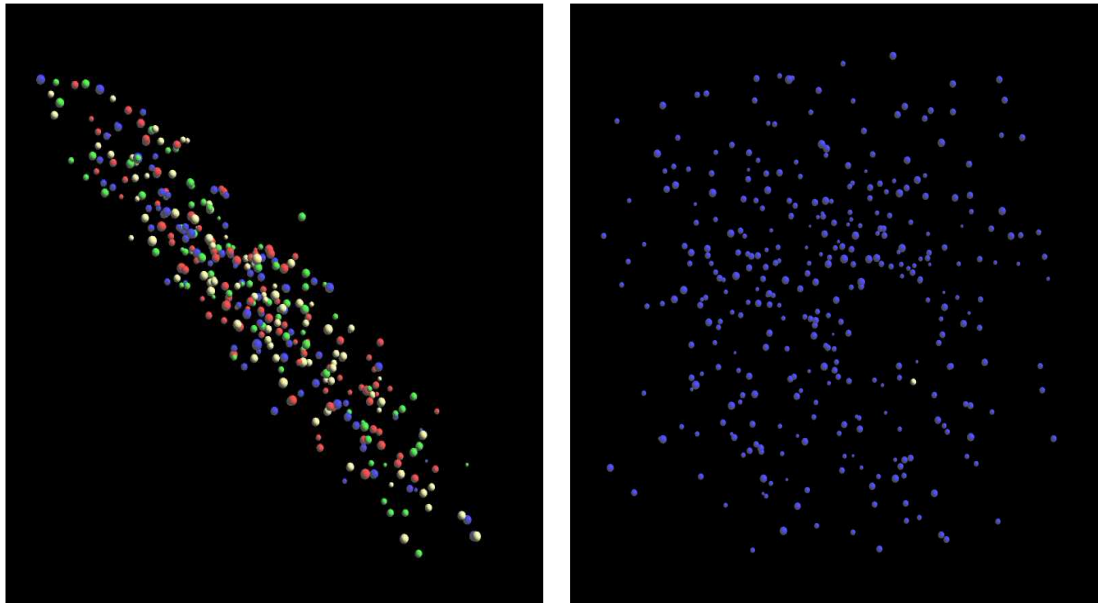


Figure 16 – stimuli used in Experiment #2. Left: side view of the dataset shown in previous figure, with the position outlier clearly visible; right: a dataset with 343 points and a colour outlier.

Participants

Participants were recruited via printed advertisement on campus and by personal communication. All participants were fully briefed and debriefed, and signed an informed consent form before commencing the experiment. They were paid for their time. Overall 20 participants (11f/9m) took part in the study over a period of two weeks. Of these, 17 were in the 21-30 age group, two in 16-20 and one in 31-40. Nine participants were students in the School of Computing and Creative Technologies; four from the School of Social and Health Sciences; two from Contemporary Sciences and five were not students of Abertay University. Handedness questionnaire classified eight of them as pure right-handed, nine as mixed right-handed, and three as mixed left-handed; there were no purely left-handed participants.

Predictions

From the experiment design, it is possible to make specific predictions from the hypotheses listed at the end of the Introduction.

1. In active rotation mode, interaction time will either linearly increase with stimulus set size or stay constant, but not decrease or show a non-linear relation.
2. There will be a difference in performance between active and passive interaction modes.
3. When the outlier is defined by colour, response time will not be affected by data set size.
4. In line with the results of Experiment #1, in the passive rotation mode, position outliers in larger data sets will take longer to detect.
5. The mouse log, when analysed to split overall response time into interaction time and motor response time, will demonstrate that data set size affects interaction time but not motor response time.

Results

Mouse log analysis

During the main experiment task, all mouse events were recorded and stored in .csv files, one file per trial, organised in folders by participant. Every row in a file stored a time-stamp, type of event, and the screen coordinates of the mouse pointer. The events were time-stamped relative to the start of the trial, with 1msec resolution and ± 10 msec precision. Distinct event types included button press or release and mouse move or drag. These logs were later processed separately for each of the three types of participant task: colour outlier trials; position outlier trials with passive rotation control; and position outlier trials with active rotation control (see Figure 17 for a summary of colour outlier trial logs). The aim of the analysis was to derive a number of variables that would give a more comprehensive picture of participants' actions during the trial.

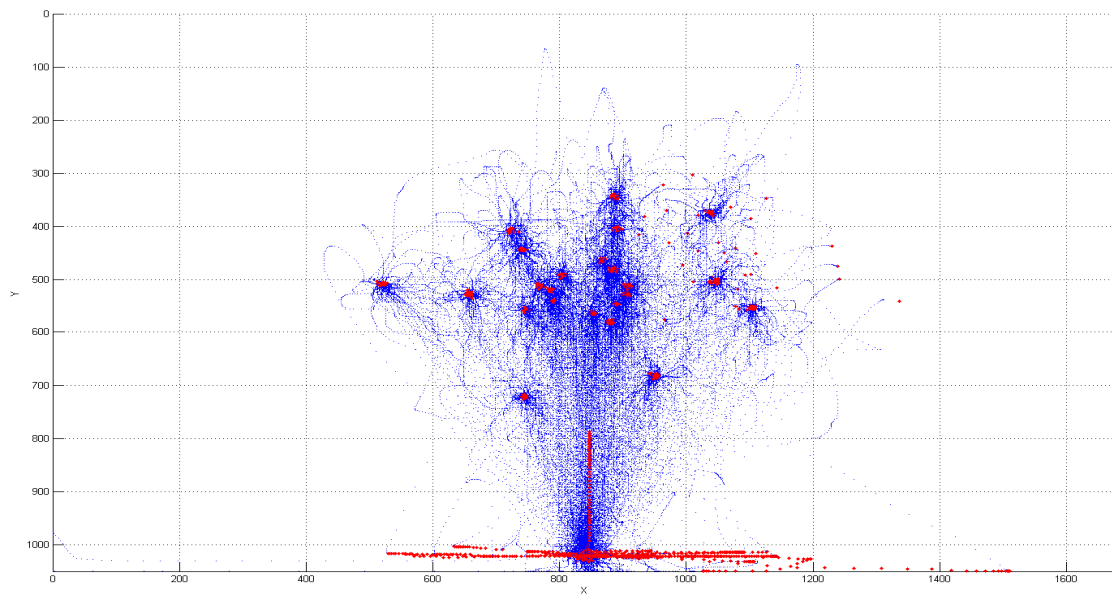
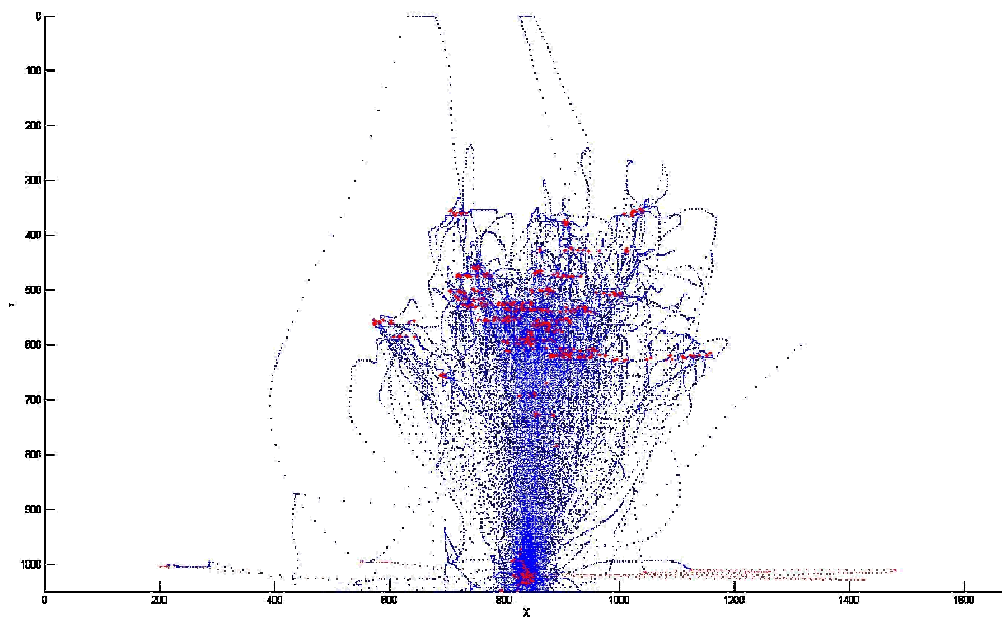


Figure 17 – mouse movements during colour outlier trials; overlay of all trials. Blue dots: mouse movements without button press; red dots: button presses and mouse movements with button pressed (dragging). The axes are screen XY coordinates in pixels.

In colour outlier cases there was no interaction with the scene view. Therefore the only use of mouse log was to identify the trials in which the participant, by mistake, started scene rotation. These cases were not numerous (less than 1%), and their inclusion or exclusion did not change the pattern of colour outlier detection (see next section).

For trials with position outlier and passive view control mode (Figure 18), a more elaborate mouse log analysis became possible. First, mouse press events were sorted into three categories: presses at the button; presses in the scatterplot area; and presses to other screen areas. The time course of the trial was then divided into three stages. The first stage was measured from the beginning of the trial (pressing on the 'next' button) until the first press on 'rotate' button. That is the pre-interaction waiting time, during which the stimulus was present on the screen, but no depth-from-motion cues were available to the viewer. The second time stage was measured from the start of first rotation until the end of last rotation. This is the time when the scatterplot is rotating and the 3D information in form of the relative motion cues was available to the viewer; this interaction time is the focus of this experiment. The third and last stage was measured from the end of last rotation until the

first click in the scatterplot area. This is the post-interaction time, or the motor response stage, i.e. reaching to the target and clicking on it. In 28% of the correctly answered trials, more than two press events in the area of the button were logged during the second stage, i.e. during the interaction time. In these cases, the time between the odd-numbered and the even-numbered presses was added up. This was the net time of the scatterplot rotation. Overall, this analysis resulted in five derived variables for each trial: waiting time; gross interaction time; net rotation time; response time; and the number of rotation episodes.



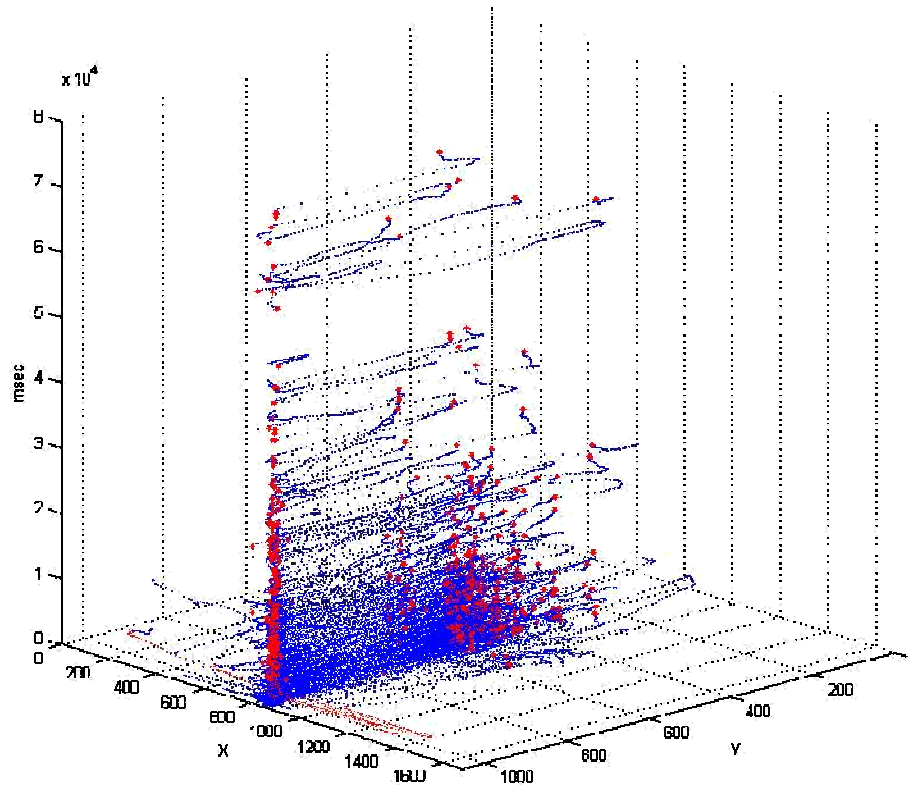


Figure 18 – mouse movements during position outlier trials with passive view control mode; overlay of all trials. Top: screen coordinates view (the axes are screen XY coordinates in pixels); bottom: XxYxTime view (X and Y axes are screen coordinates in pixels; Z axis in time from the start of the trial in milliseconds) Blue dots: mouse movements without button press; red dots: button presses and mouse movements with button pressed (dragging).

In active rotation trials (Figure 19), every trial time course was first divided into the same three stages: waiting time, interaction time and response time. Then, the interaction stage was divided into separate side-to-side mouse sweep events by analysing¹⁸ the function $y=f(t)$ for local minima and maxima. This resulted in the following derived variables for each trial: number of sweeps, average sweep time, average sweep range (in degrees), and a derived variable of average sweep speed. From the wealth of data provided by the mouse logs it is possible to compute other derived variables: button-to-target path lengths, number and timing of missed target clicks etc. These were not calculated since they bear no relevance to the questions asked in this study.

¹⁸ Minimax search was performed with a freeware Matlab function *peakdet*, written by Eli Billauer (explicitly not copyrighted)

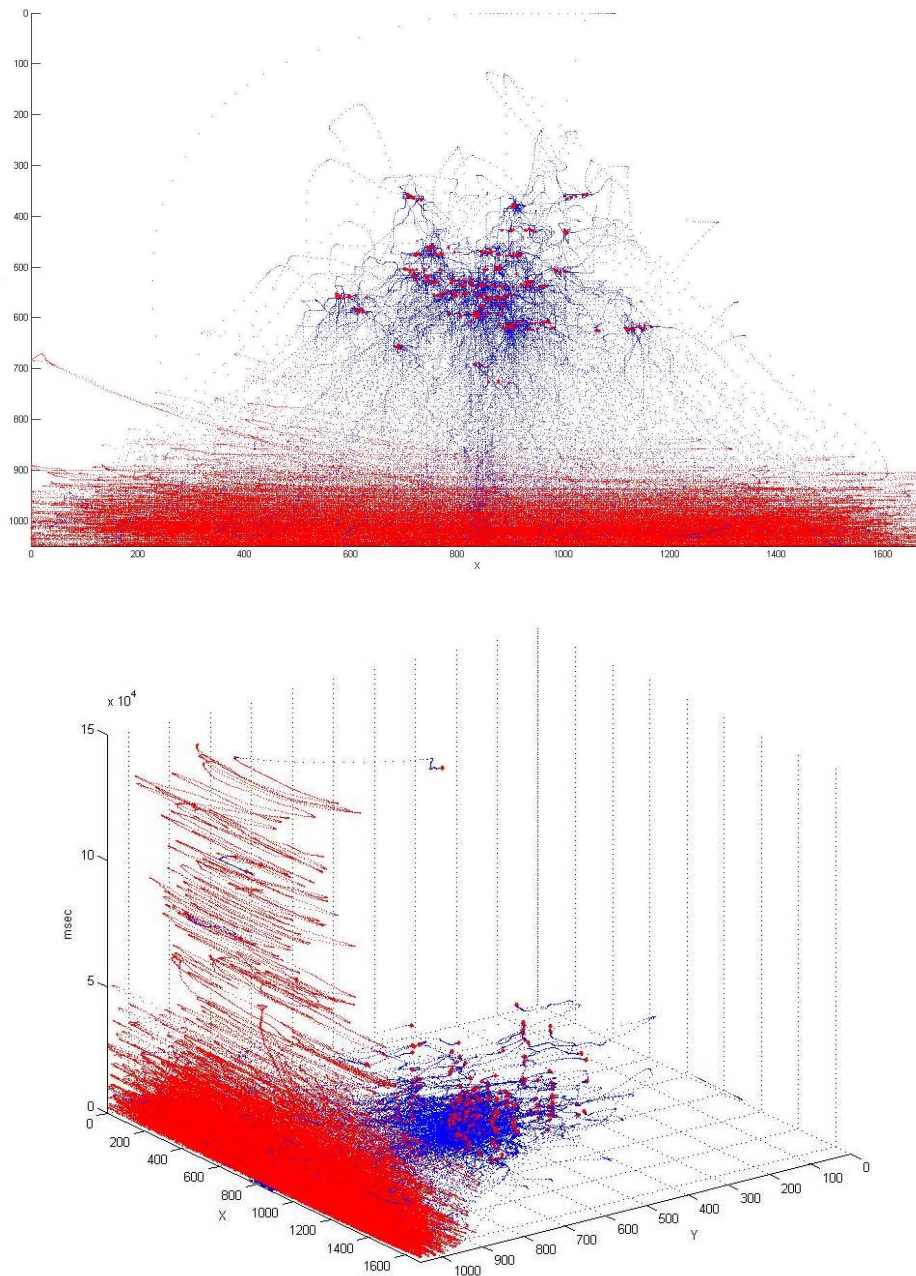


Figure 19 – mouse movements during position outlier trials with active view control mode; overlay of all trials. Top: screen coordinates view (the axes are screen XY coordinates in pixels); bottom: XxYxTime view (X and Y axes are screen coordinates in pixels; Z axis in time from the start of the trial in milliseconds) Blue dots: mouse movements without button press; red dots: button presses and mouse movements with button pressed (dragging).

Data screening

Overall 2560 separate trials were considered for analysis. First, accuracy rates were tested separately for each combination of data set size and trial type (colour outliers, passive or active rotation) (see Figure 20). The accuracy levels were far above chance level and did not vary much between data set sizes; in all subsequent analyses, trials with incorrect answers

were discarded and analyses were made on correct answer trials only. In the next step, each of the three time components identified by mouse log analysis (wait time, interaction time and response time) were tested for distribution normality, again, separately for colour outliers (not shown) and for passive and active rotation modes of position outliers (see Figure 21). Analysis showed that all time data in this experiment is distributed log-normally; therefore in all subsequent time analyses, a natural logarithm link function was used.

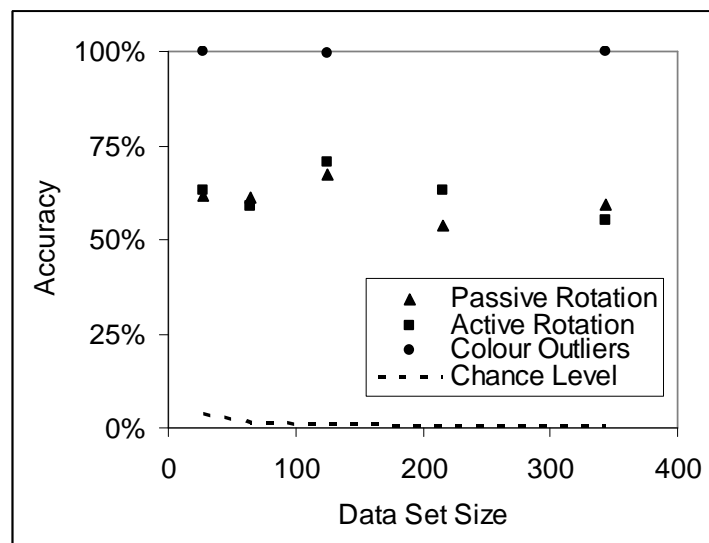


Figure 20 – Percent of correct answers. Chance level is the inverse of the number of points.

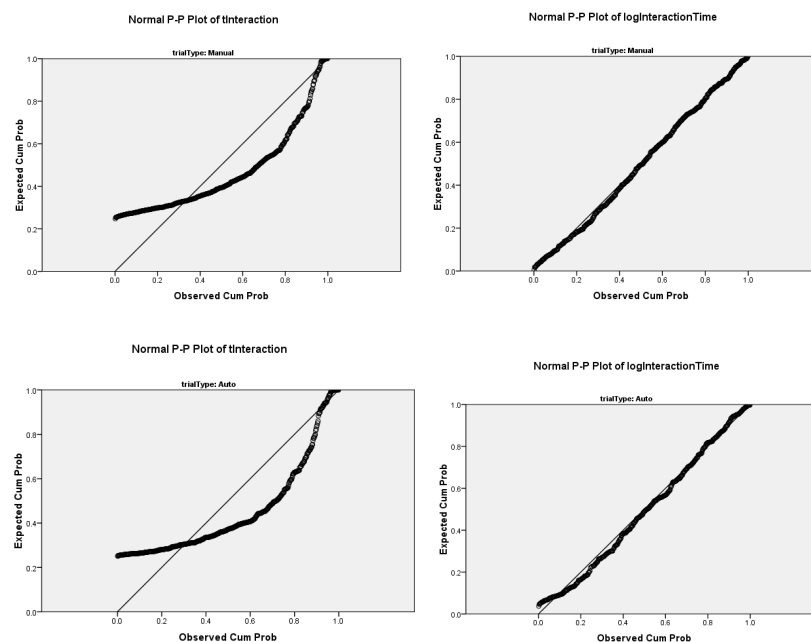


Figure 21 – P-P plots of interaction time. Top: passive rotation; bottom: active rotation; left: no transformation, right: natural logarithm link function.

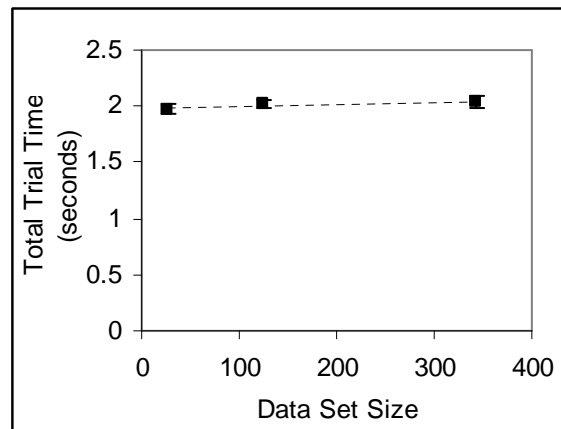


Figure 22 – effect of data set size on total response time in trials with colour outlier

Regression analysis of time versus data set size was run next with the model $\ln(\text{Time}) = b_0 + b_1 * \text{DataSetSize}$, separately for each trial type, and, for position outliers, separately for pre- and post-interaction time component (interaction time analysis is the main focus of this experiment and will be addressed in the next section). For colour outlier trials, the analysis does not show an effect: $b_1 < 0.001$, $\beta = 0.036$, $p = 0.263$ (see Figure 22). For position outlier trials, the only significant effect is the negative effect on pre-interaction (wait) time in passive rotation cases (see Figure 23 and Table 1).

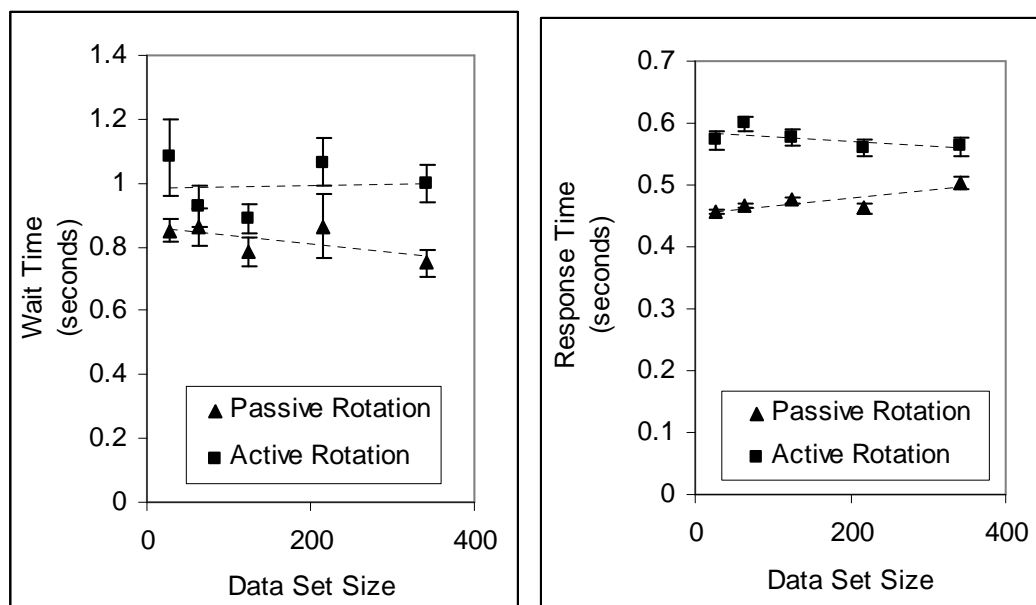


Figure 23 – effects of data set size on pre- and post-interaction time. Left: pre-interaction (wait) time; Right: post-interaction (response) time. Error bars indicate ± 1 standard error of mean.

	Wait Time	Interaction Time	Response Time
Passive Rotation	$b_1 < 0.001$ $\beta = -0.123$ $p = 0.007$	$b_1 = 0.002$ $\beta = 0.169$ $p < 0.001$	$b_1 < 0.001$ $\beta = 0.025$ $p = 0.582$
Active Rotation	$b_1 < 0.001$ $\beta = -0.001$ $p = 0.974$	$b_1 = 0.002$ $\beta = 0.189$ $p < 0.001$	$b_1 < 0.001$ $\beta = 0.052$ $p = 0.244$

Table 1 – results of regression analysis of wait, interaction and response time as a function of data set size, separately for active and passive rotation modes. Only trials with position outlier, and only trials where participants gave the correct answer were analysed.

Main analysis

The effect of data set size on interaction time (see Figure 24) was determined using a linear regression model with a logarithmic link function: $\ln(\text{time}) = b_0 + b_1 * \text{DataSetSize}$. A significant effect was shown in both passive and active rotation modes (see middle column in Table 1).

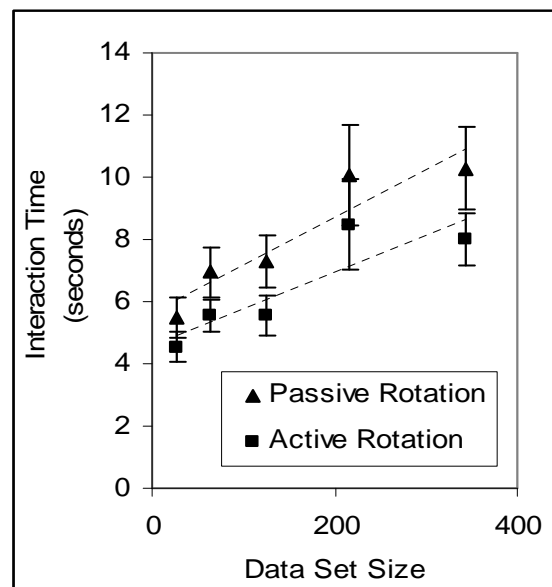


Figure 24 – effects of data set size on interaction time. Error bars indicate ± 1 standard error of mean.

A simple comparison of interaction times between active and passive rotation modes, with the same participant and for the same dataset shows that interaction with active control is on average 1.33 seconds faster. However, in a paired t-test with logarithmical transform of the interaction time, the difference does not show as statistically significant: $t(377) = 1.853$; $p = 0.065$.

Additional analyses

In order to account for between-subject and between-trial variability in the analysis of interaction time, the experimental data included a number of independent variables: parameters of the trial design, such as the slant of the reference plane and trial presentation order; numerous performance variables measured during trial, such as the mouse movement during active interactivity trials; and several participant profile variables such as sex or age.

The first set of additional analyses looked at the effects of different trial parameters: reference plane slant, trial presentation order and the colour of the target sphere. First, the effect of reference plane slant ('floor' or 'ceiling') on interaction time was assessed using a full-factorial GLM, with the dependent variable being the natural logarithm of interaction time. Only position outlier trials with correct answers were analysed. The independent variables were data set size and reference plane slant. This analysis was run separately for two interactivity modes. Reference plane slant emerged as a highly significant factor for both passive ($F(1,476)=38.273$; $p<0.001$) and active ($F(1,487)=16.895$; $p<0.001$) interactivity modes (see Figure 25A). No significant interaction between data set size and reference plane slant was found ($F(4,476)=0.662$; $p=0.618$ for passive and $F(4,487)=0.909$; $p=0.458$ for active interaction mode). A GLM analysis regarding the colour of a target sphere found no effect of colour on interaction time ($F(3, 482) = 1.333$; $p=0.263$ for passive and $F(3,493)=0.878$; $p=0.452$ for active interaction mode; see Figure 25B).

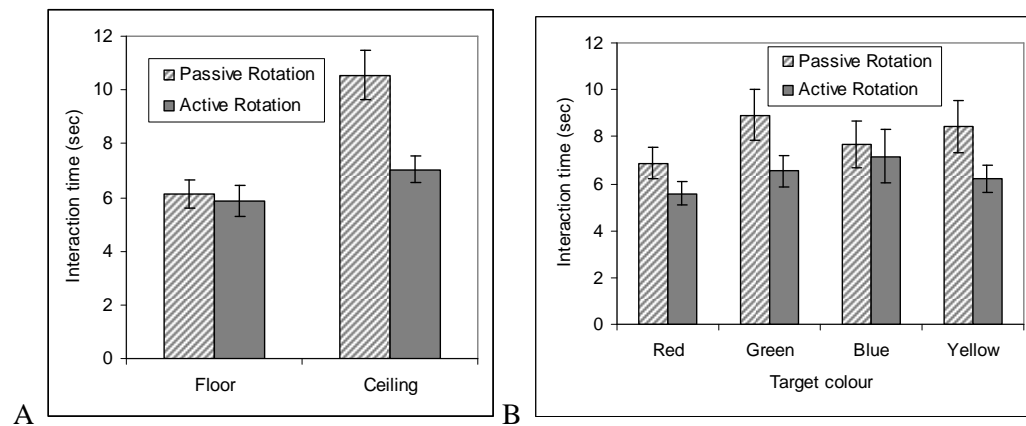


Figure 25 – A: effects of reference plane slant on interaction time. Left: ‘floor’ slant (bottom closer to the viewer); right: ‘ceiling’ slant (top closer to the viewer). B: effects of target colour on interaction time. Error bars indicate ± 1 standard error of mean.

A coarse, block-wise analysis of task learning in the course of the experiment (see Figure 26) shows a significant decrease in interaction time from block to block for active rotation ($F(3,493)=3.239$ $p=0.022$). The difference between blocks for passive rotation was also found to be significant ($F(3,482)=5.707$ $p=0.001$) but not with the blocks in the order of presentation, but because performance in block 3 differed significantly from the rest. No interaction of block presentation order with data set size was found. This is supported by a higher-resolution analysis assessing the linear regression of log interaction time as a function of data set size and order of presentation of every trial, separately for active and passive interaction modes (Table 2).

	Data Set Size	Trial Order of Presentation
Passive Rotation	$b = 0.002$ $\beta = 0.171$ $p < 0.001$	$b = -0.001$ $\beta = -0.051$ $p = 0.260$
Active Rotation	$b < 0.002$ $\beta = 0.185$ $p < 0.001$	$b = -0.003$ $\beta = -0.139$ $p = 0.002$

Table 2 – regression analysis of training effects

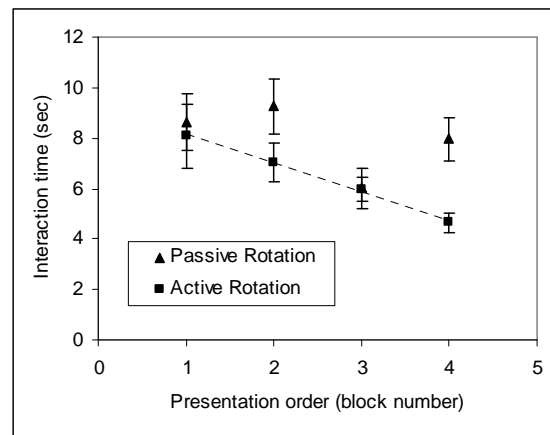


Figure 26 – effects of order of presentation on interaction time. X-axis: trial blocks in order of presentation. Dotted line: trend for active rotation (the trend for passive rotation was not significant). Error bars indicate ± 1 standard error of mean.

To summarise, reference plane slant and trial order, but not target colour, have an effect on interaction time. Introducing reference plane slant and trial order into main analysis offers a more accurate account of the effect of data set size on interaction time. An enhanced linear regression analysis was performed separately for active and passive interactivity mode, with slant and trial order variables being entered stepwise. In both interactivity modes, the resultant model of the stepwise regression was the full model $\ln(\text{time}) = b_0 + b_1 * \text{DataSetSize} + b_2 * \text{Slant} + b_3 * \text{TrialNumber}$ ($R^2 = 0.112$ for passive and 0.074 for active rotation mode; see Table 3 for coefficients).

	Predictor	b	β	p
Passive rotation				
b_1	Data set size	0.002	0.174	<0.001
b_2	Slant	0.566	0.275	<0.001
b_3	Trial number	-0.011	-0.103	0.017
Active Rotation				
b_1	Data set size	.002	.184	<0.001
b_2	Slant	.319	.172	<0.001
b_3	Trial number	-.009	-.085	0.050

Table 3 – Experiment #2 enhanced main analysis, linear regression coefficients (b slope, β standardized slope and significance). ‘Floor’ slant was encoded as 1 and ‘ceiling’ slant as 2.

A pairwise contrast of interaction time in passive and active rotation modes shows different patterns for different reference plane slants. With ‘floor’ slant, the active rotation is on average 0.264 seconds slower than the passive ($p=0.779$, non-significant), whereas with ‘ceiling’ slant, active rotation is 4.23 seconds faster, and that difference is statistically significant ($p=0.01$).

The next set of analyses investigates the effects of participant profile on performance. Participant profiles comprise record of sex, age group, School of study, handedness and digit span. Of these, age group data is not varied enough, with only three participants outside the 21-30 age bracket, and is therefore excluded from the analysis. Basic per-participant performance indicators are speed and accuracy, separately for each of the two interactivity modes. ANOVA analysis found no effect of sex, handedness or School of study on either speed or accuracy (p -values ranging from 0.24 to 0.94). A regression analysis of speed and accuracy as a function of digit span, separately for interactivity modes, did not find any pattern either (p -values ranging from 0.3 to 0.84).

The per-participant performance indicator that is most relevant to the questions set in the main analysis is the effect that stimulus data set size has on interaction time, separately for interactivity modes. This performance indicator is b_1 in the function $\ln(\text{time})=b_0+b_1*\text{dataset size}+b_2*\text{slant}$. Regression analysis of this model is not statistically powerful on a per-participant basis. However, even though the regression model is not significant, slope coefficient on its own is still a valid derived variable. This regression analysis yielded six derived variables per participant: mean interaction time; effect of slant; and effect of data set size, all these separately for each interactivity mode. These variables were compared with participant profile variables (sex, School, handedness and digit span) in a series of MANOVA analyses. None of the participant profile variables had a significant effect on the interaction time profile (p -values all higher than 0.3). It is thus possible to conclude that personal profile has no effect on aspects of performance that relate to the effect of slant and data set size on interaction time.

A performance indicator that emerged in the analysis of effects of trial presentation order is the amount of performance improvement during the experiment. Per-participant analysis of effects of trial presentation for active interactivity only was run (see Figure 27). Three out of twenty participants (LDN, JJ and DG) showed improvement in speed over the course of the experiment that was significantly better than the improvement of other participants ($t < 0.01$). All three were right-handed, studied in the school of Computing and Creative Technologies, and had a digit span score that was also significantly higher than that of the rest ($t = 0.04$).

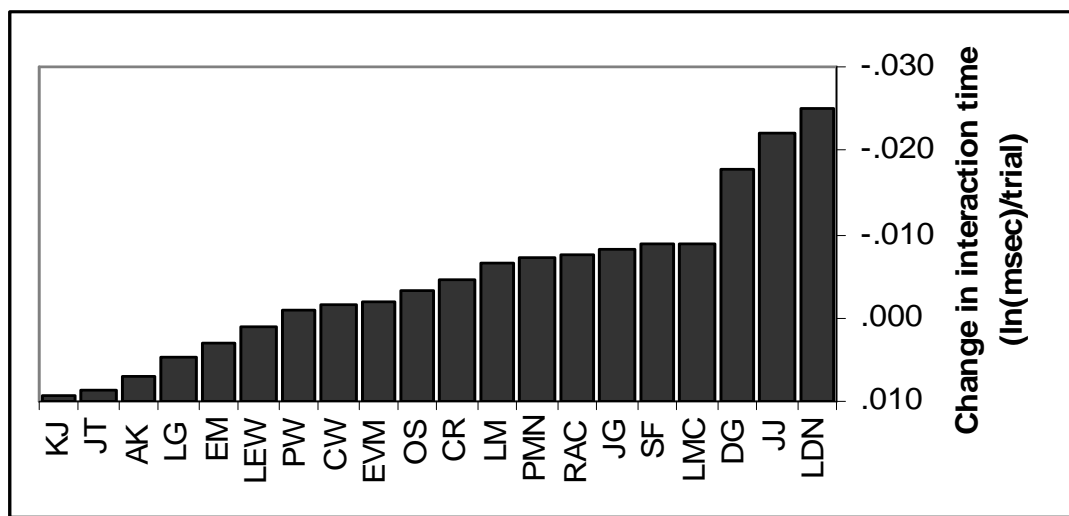


Figure 27 – change in interaction time over the course of Experiment #2, for active interactivity trials only.

The last set of additional analysis consists of analysis of per-trial secondary performance indicators that emerged from mouse log analysis of active interactivity trials. These indicators reflect planned behaviour of participants during interaction and thus can be seen as indicators of participants' strategy. Mouse analysis resulted in division of interaction into a series of 'sweeps', i.e. left or right mouse movements that rotated the scene. For each such sweep, range and time were calculated. These have been processed and averaged to produce the average sweep range and the average sweep speed for each trial. These two indicators were chosen because they do not directly affect the overall interaction time.

Correlation analysis showed that trial interaction time is positively correlated with average sweep range (Pearson $r=0.47$; $p<0.001$) but not with average sweep speed ($p=0.2$). Overall comparison of average sweep range and speed between trials with correct and incorrect answers shows that in trials with correct answers, sweep range was significantly smaller ($p<0.001$) and sweep speed higher ($p<0.01$).

Per-participant regression analysis of effect of trial presentation order on average sweep range and average sweep speed resulted in estimates of personal strategy changes during the course of the experiment. These changes in strategy were then correlated to change in performance (specifically changes in interactivity time). A significant correlation of change in interactivity time and change in range was found (Pearson $r=0.7$; $p<0.001$). The correlation of change in interactivity time and change in speed was not significant ($p=0.6$). Those participants who reduced their sweep range performed better. No effect of change of sweep speed on change in interaction time was found (see Figure 28).

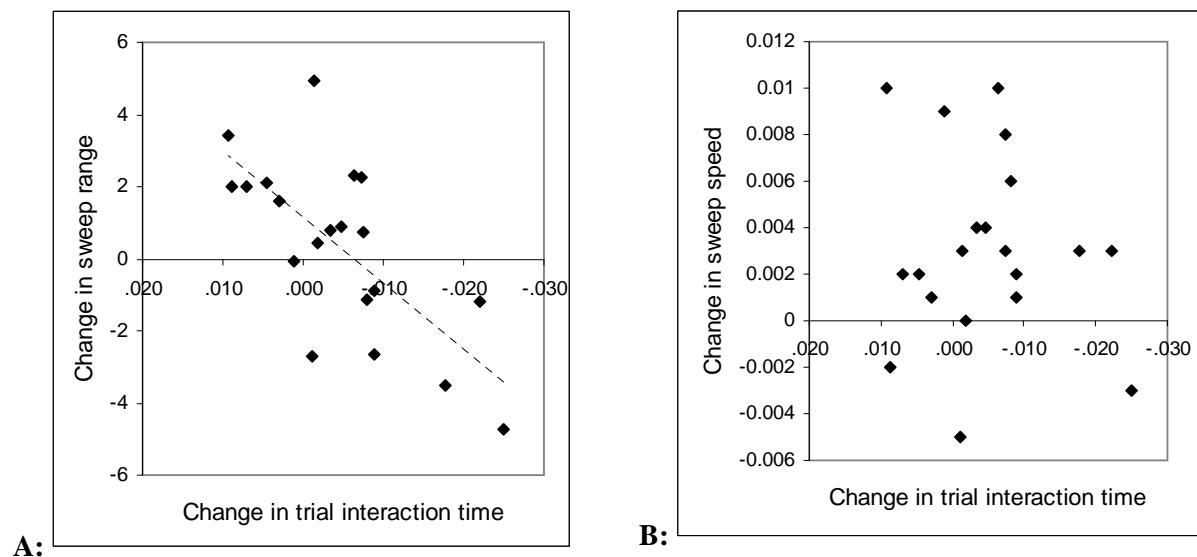


Figure 28 – juxtaposition of personal changes in strategy and changes in performance. A: changes in sweep range; B: changes in sweep speed. Dotted line in A signifies linear trend.

Discussion

Analyses have shown that in trials with outliers defined by colours, stimulus array size does not affect response time. Detailed mouse log analysis has shown that in trials with outliers defined by position, stimulus array size did not affect waiting time or motor response time.

The main analysis has shown a strong effect of stimulus array size on scene rotation (interaction) time. This effect was observed in trials with both active and passive interactivity modes. The difference in interaction time pattern between trials with different interactivity modes was found to be barely noticeable; still, active interaction was, on the average, slightly faster. When the results were analysed separately for each slant, however, the difference did not show in floor slant, but was clear in ceiling slant. This is in agreement with previous research: Scott-Brown (1996) found a slight tendency for non-veridical speed perception for horizontally translating lines. Participants showed a slight reduction in perceived speed for lines translating above fixation. Much less bias was found for identical line translation below the fixation point.

Additional analyses uncovered two unexpected factors that significantly affected interaction time in trials with position outliers: reference plane slant and trial presentation order. Analysis of reference plane slant has shown that trials with a ‘ceiling’ slant (with top edge being closer to the viewer than the bottom edge) were noticeably more difficult to resolve. This was reflected both in longer interaction times and higher error rates. Trial presentation order affected interaction times as well: participants’ speed noticeably improved over the course of the experiment, but only when the participants had active control over the scene rotation. This reflects changes in strategy (whether conscious or unconscious) during the course of the experiment. An in-depth analysis of strategy indicators allows for tentative attribution of this improvement in overall speed to changes in scene rotation pattern. In particular, making shorter separate rotation movements led to shorter overall interaction times.

Some participants have improved in this task more than the others. However, no conclusive explanations regarding the factors that set these participants apart from the rest were possible given the available data. Specifically, participant profile factors that were collected, such as participants’ gender, handedness or memory span, did not have any observable effect on either performance or improvement in performance in the task. Partly that could be

attributed to relatively unvaried and unbalanced sample. For example, no big differences in age were present in the sample. A possible indication of the effect of memory span on learning capacity was seen in that the group of three participants with significantly higher performance improvement rates also had digit span results that were significantly higher than the results of the rest of the participants.

During the experiment design stage, several hypotheses were proposed and corresponding predictions formulated. The empirical findings have resolved most of these.

1. The experiment produced results interpretable within the ‘pop-out’ paradigm when used in an interactive task: pattern of position outliers’ detection was consistent with the theoretical parameters of serial visual search processes.
2. Interactivity changed performance pattern, however in an unexpected way. Instead of making the target pop out or changing the parameters of the visual search (slope or constant shift), as hypothesised, active interactivity enabled learning: improvement in performance over the course of the trial.
3. Performance consistent with ‘pop-out’ effect was observed when the outlier was defined by colour. Colour outliers were detected immediately, consistent with previous research.
4. A strong difference in performance between trials with different reference plane slants clearly shows that the 3D scatterplot outlier detection task is based on 3D scene reconstruction rather than on local motion comparison, since the local motion around the outlier is not affected by overall plane slant.
5. Number of elements affected the time required to identify the outlier, rather than the time required to make a motor response. This supports attributing changes in performance patterns to perceptual processes, rather than to motion planning.

The last two points explicitly address criticisms regarding Experiment #1, and these findings can be applied retroactively to analysis of Experiment #1 results. Specifically, difference

between reference plane slants supports the involvement of global analysis; and mouse movement analysis clearly shows that stimulus set size affected interaction time only.

Conclusions and implications

On a methodological level, the results show that ‘pop-out’ methodology can be applied to complex interactive tasks; also that with some in-depth analysis, the stages of interaction can be easily differentiated, and interactive strategy indicators gathered and analysed. On the applied level, the benefits of active interactivity were demonstrated in that allowing adjustment of presentation parameters enabled learning and led to improved performance. Not all participants benefitted from this in equal measure; further research into what makes some participants improve could yield relevant results.

From the point of the dichotomy between immediate and serial perception, it seems that even improved performance did not seem to exhibit patterns of immediate perception, i.e. ‘pop-out effect’. This was most clear in the time differences between colour and position outlier trials with more than 300 points. Colour outliers were detected in about 2 seconds, while position outliers took more than 8 seconds. This further supports the conclusion that emerged from Experiment #1: that 3D scatterplot is a suboptimal visualisation technique for outlier detection.

Chapter 8. Experiment 3: roles of high and low-level objects in trend detection

Introduction¹⁹

Experiments #1 and #2 explored interactive information comprehension in a 3D scatterplot, in light of the Reverse Hierarchy Theory. The results suggest that this theory provides a feasible framework for the assessment of processes involved in visualisation comprehension. Both of these experiments focused on the same visualisation task: detection of a single outlier. While outlier detection is a common Visual Analytics task, it is by no means the only one, and studying solely outlier detection reduces the generalisability of experimental findings. Experiment #0 indicated that at least two other tasks are commonly used in Visual Analytics: cluster detection and trend detection. An experiment focusing on cluster detection in a 3D scatterplot, with clusters defined by 3D displacement, would be relatively similar, task-wise, to Experiments #1 and #2. An experiment in trend detection, on the other hand, would be less similar and thus more likely to contribute to broader outlook on the whole field, complementing the experimental programme of this thesis.

In research, detection of trends in data, i.e. structured relationships between variables, is at the core of data analysis, in the same manner as outlier detection is an indispensable part of preliminary data screening (Box, Hunter, & Hunter, 1978; Peck, Olsen, & Devore, 2008). Statistical analyses for linear trend detection, such as the linear regression, are the workhorse of a researcher (ibid.). However, since these analyses operate in terms of linear trends only and do not allow identifying non-linear relationships, visual exploratory analysis is an indispensable step in data analysis (Anscombe, 1973; Tukey, 1977). Such visual analysis is

¹⁹ The results of this experiment were presented at Applied Vision Association Easter Meeting 2010 (see Appendices and the attached CD for details)

often performed using 2D scatterplots (Tabachnik & Fidell, 2006); in the same manner, 3D scatterplots can be expected to be useful for detection of non-linear tri-variate trends.

In a similar way to the outlier detection task designed for previous experiments, a trend detection task should consist of identifying a non-random positioning of points in a 3D scatterplot. A stimulus scene should thus comprise of several groups of points, one being the target group and all others being distracters. In the target group there should be a non-random (i.e. structured) relationship between its constituent points' positions. To ensure that the task is not reducible to planar trend detection, this relationship should be made inherently three-dimensional. Points in the distracter groups should be positioned randomly. The groups should be presented superimposed in space, and distinguished by some non-spatial parameter such as colour or shape.

Application of the 'pop-out' methodology (Chapter 5) in such a task should consist of modulating the number of objects in a scene and measuring the time to identify the target group, i.e. a non-random positioning of points. This design raises an unresolved theoretical question: what is an object in this task? A prediction based on the Feature Integration Theory (Treisman & Gelade, 1980) would be that since, as shown in Experiments #1 and #2, the 3D position of a single point is a conjunction of low-level features, every point is a complex perceptual object in its own right. Conversely, a prediction based on the Reverse Hierarchy Theory would be that the object in this task is a highest-level task-relevant arrangement, i.e. a group of points of the same colour or shape.

The empirical prediction following from the Feature Integration Theory would be that increasing the number of points increases the response time. Reverse Hierarchy Theory, on the other hand, would predict that increasing the number of point groups would increase the difficulty of the task, and would make no prediction of the effect of increasing the number of points. Since the Reverse Hierarchy Theory postulates the tie between 'pop-out' effect and formation of high-level semantics, which is a crucial assumption in the proposed use of the

‘pop-out’ methodology as a visualisation comprehension evaluation tool, an experiment was planned to test these two predictions.

From a technical perspective, Experiment #3 is built as an extension of Experiment #2, with the same visualisation and manipulation options. The difference between experiments is in the task, which is reflected in different stimulus sets and response options. Stimulus sets are combinations of a variable number of point groups, with a variable number of points in each group. Each of these two variables is manipulated independently between trials. Since Experiment #2 has demonstrated an immediate perception of colour, the point groups in the scatterplot are colour-coded, allowing the participants to respond by indicating the colour of the target group. Also building on the results of Experiment #2, participants are given the option to interactively rotate the scene, using the same interface as in the active rotation condition of Experiment #2.

To summarise, Feature Integration Theory and Reverse Hierarchy Theory predict different patterns of behaviour for trend detection in 3D scatterplots. Based on the evidence of visual search in 3D scene perception that was gathered in Experiments #1 and #2, both theories predict that trends in a 3D scatterplot will not be immediately perceived. However, the pattern of predicted visual search is different between the two theories. Feature Integration Theory predicts an increase in response time with an increased in *overall* number of points in a scene; Reverse Hierarchy Theory predicts an increase in response time with an increase in number of point *groups*. Experiment #3 is designed to test these predictions. The design used is a within-subject design with five levels of number of points per group, three levels of number of groups, and two levels of depth displacement noise added to the target group. The measured variable is the net interaction time for correct answers.

Methods

Apparatus

The apparatus in this experiment was the same as in Experiment #2, with several minor alterations. Keyboard keys 1, 2 3 and 4 (top row of a standard QWERTY keyboard) had coloured stickers added to them: red, green, blue and yellow respectively; these were used for making the response in the main task. The experiment was run in a different experimental chamber, but one that shared the same dimensions and lighting conditions as the previous laboratory. Lastly, an office chair without wheels was used instead of the office chair with wheels, to ensure constant seating position of the participant.

Procedure

At the beginning of the experiment, participants were given an information leaflet and a consent form. The information leaflet explained that the task in each trial was to identify the colour of the group of points that were positioned non-randomly. After signing the consent form, participants filled in the details in the main dialog form: initials, age, gender (Male or Female) and school of study: Social and Health Sciences (SHS), Computing and Creative Technologies (CCT), Contemporary Sciences (CS), Dundee Business School (DBS) or Other. After filling in the details, the participants undertook a computerised version of digit span test, the same as the one used in Experiment #2.

Before starting on the main experiment task, participants undertook a short training exercise to familiarise themselves with the task. The training comprised nine trials of increasing complexity. The first training trial was to identify a square grid structure with a single distracting noise group. More noise groups and more complex 3D structures were gradually added, with the last three trials being of the same structure as the main task.

The experiment itself comprised 120 self-paced trials presented in a random order. After every 30 trials, a pause screen was displayed for at least 30 seconds. The pause screen was

gray, with black text in the middle. For the first 30 seconds, the text read “Block X/4. Please rest your eyes for at least 30 seconds”, where X was either “1”, “2”, or “3”. During these 30 seconds, no input from the keyboard or mouse was processed. After 30 seconds, the text changed to “Please press any key to continue” and any input from the keyboard would close the pause screen and advance the main test to the next trial. During the briefing, participants were specifically encouraged to take a break and to rest their eyes between the sets. The duration of the experiment, including explanations and training, was approximately 30 minutes per participant.

The user interface in this experiment was identical to the active rotation mode interface in Experiment #2, with the exception of the response option. To finish the trial, instead of clicking on the target sphere, participants were instructed to press one of the four colour-coded buttons (keys 1, 2, 3 and 4 on the top row of keyboard). This response was time-stamped and recorded together with all the mouse movements and actions.

Stimuli

In contrast to Experiment #1 and #2, where in every trial a single data set was shown, in Experiment #3 the stimulus in a trial consisted of a composition of either two, three or four datasets displayed simultaneously. The datasets were presented in the same space in the centre of the screen, but in different colours. One was the target data set, and the rest were distracter data sets. In each trial, all sets had the same number of points. In different trials the number of points in the sets varied: 64, 100, 144, 196 or 256 points²⁰. The manipulation of the number of datasets and number of points per set was fully balanced in the experimental design, resulting in 15 options for total number of data points in a stimulus, from 128 (2x64) to 1024 (4x256).

²⁰ $(2n)^2$ for n in the range 4 to 8

In distracter datasets, the spatial position of each point was randomly drawn from a uniform distribution, in a ± 1 range. Four sets were pre-generated for each data set size, resulting in 20 distracter data sets overall. In target datasets, the planar position of points (X and Y) was randomly generated in the same way as in the distracter sets. The depth position of each point (Z) was defined by a formula $z = a \cdot \max(x^4 - bx^2, y^4 - by^2)$. Coefficient a was either 1 or -1; coefficient b was either $\frac{1}{2}$ or $\frac{3}{4}$, resulting in 4 different target structures (see Figure 29). The resulting set of depth coordinates was stretched to fill the ± 1 range. After that, in half of the sets a random noise factor was added to the depth position. The noise factor was drawn from a uniform distribution in a ± 0.1 range. After that the alignment into ± 1 range was repeated. The combination of two a values, two b values and two noise conditions gave eight different types of target data sets, which, combined with five data set sizes resulted in 40 different target sets overall.

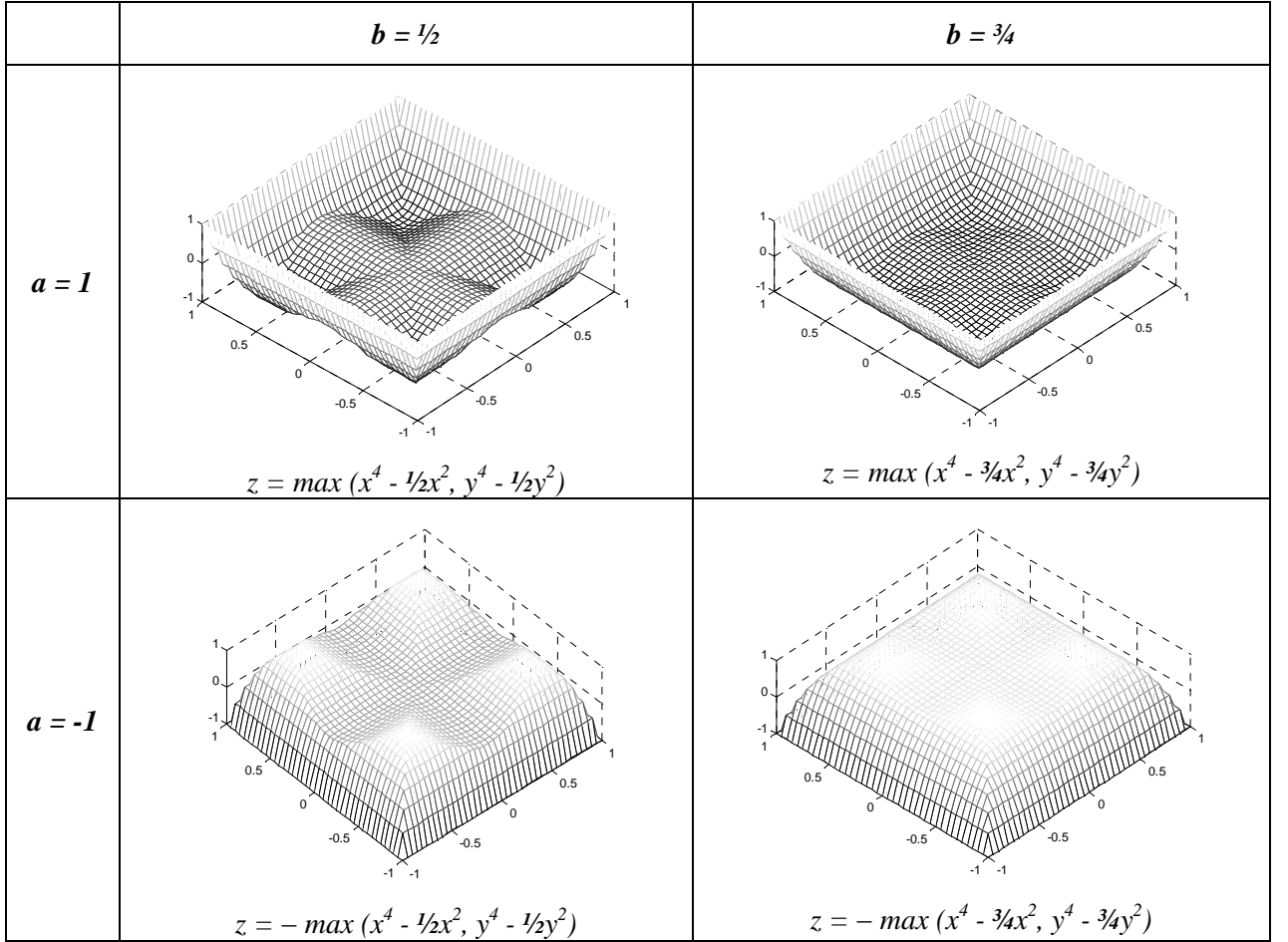


Figure 29 - 3D structures used in Experiment #3.

The size of each point was randomly assigned from a uniform distribution, ranging between 1mm and 4mm if seen at zero depth, the same size range as the one used in Experiment #2. In contrast to Experiments #1 and #2, where colour of points was predefined as well, in Experiment #3 the colour of the data sets, as well as the choice of specific distracter datasets, was random for each run, to reduce the bias caused by specific colours or distracter datasets. The colours were those used in the previous experiments: system-default red, green, blue and yellow. Sample resulting stimuli (in frontal projection) are shown in Figure 30.

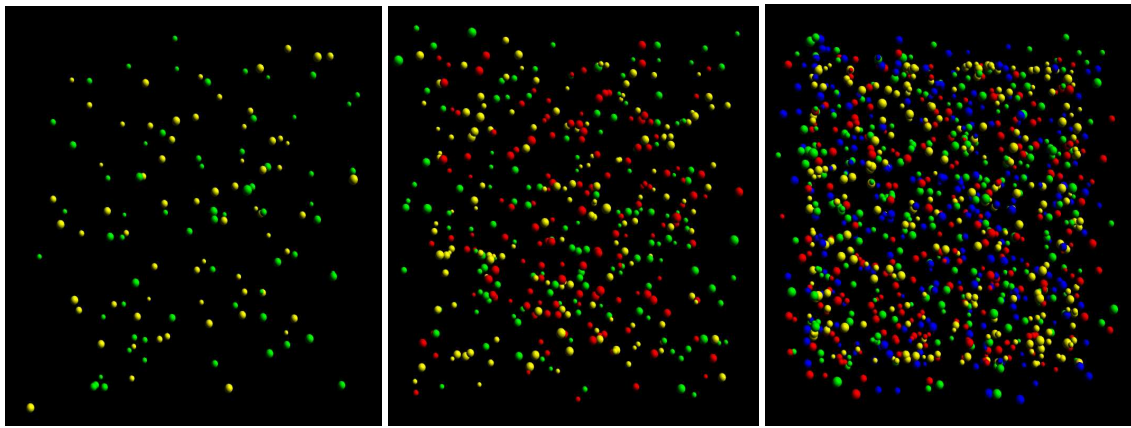


Figure 30 – sample Experiment #3 stimuli. Left panel: two datasets of 64 points. Middle panel: three datasets of 144 points. Right panel: four datasets of 256 points.

Mouse log analysis

As in Experiment #2, all mouse and keyboard responses in Experiment #3 were time-stamped and logged, and the logs processed to produce more detailed parsing of trial time course. Trial time was divided into three stages: before the first rotation; from the first to the last rotation; and from the last rotation until the response, i.e. the keyboard press. Rotation time was further refined by taking into account only the time spent rotating the scene, i.e. dragging the button icon along the length of the slider bar, and discarding the times when no rotation occurred (see Figure 31). The sum of dragging periods resulted in a measure of net rotation time. As in Experiment #2, mouse movements were parsed into side-to-side sweeps, and for each sweep, time and range were calculated. This resulted in two derived variables for each trial: average rotation range and average rotation speed.

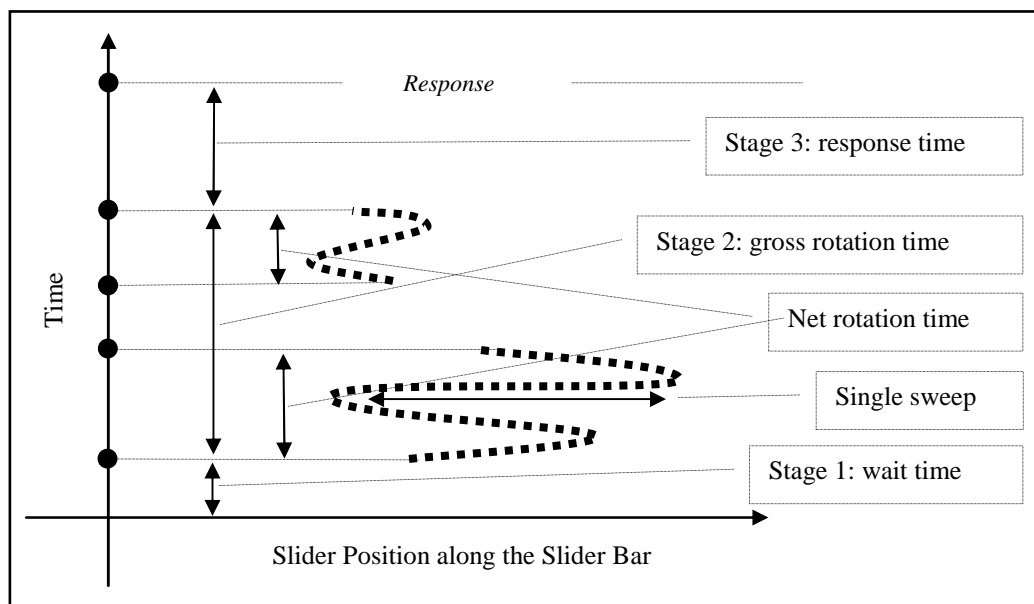


Figure 31 – schematic outline of mouse log parsing algorithm. Bold dotted line: mouse movements; solid double-arrow lines: meaningful time segments.

Participants

Participants were recruited via online advertisement on the University web portal. All participants were fully briefed and debriefed, and signed an informed consent form before commencing the experiment. They were not paid for their time. Overall 15 participants (11m/4f) took part in the study over a period of three weeks. Age ranged from 17 to 41 years, with median at 20. Nine participants were students in the School of Computing and Creative Technologies, three from School of Social and Health Sciences, one from the Dundee Business School and two were members of staff of the Abertay University.

Results

Data screening

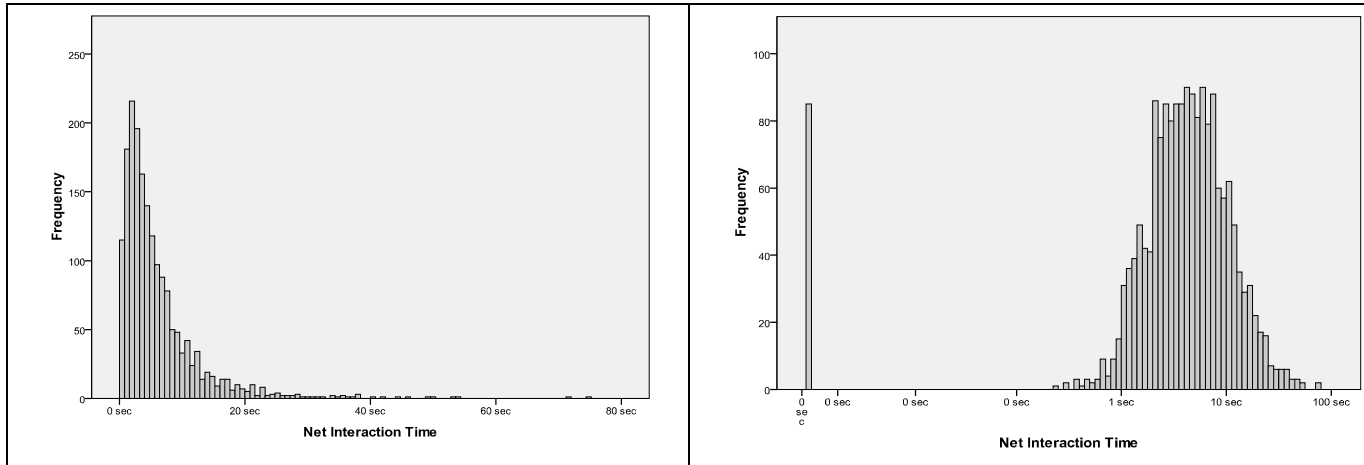


Figure 32 – histograms of net rotation time distribution. Left panel depicts data with no link function; right panel indicates data after applying a natural logarithm link function.

A histogram of all net rotation times (Figure 32 left) time clearly shows a non-normal distribution (see Chapter 5). Application of a natural logarithm link function (Figure 32 right) shows a distribution that is closer to binomial; therefore a logarithm link function will be used in subsequent response time analyses. Importantly, the second distribution shows a clear outlier group: 85 trials with zero interaction time. In these trials, participants responded without changing the view angle at all. In 66 out of these 85 trials, participants were able to correctly identify the target.

The distribution of these correct answers between different target structures is shown in Table 4, and the distribution between participants in Figure 33. These illustrate that correct answers were immediately visible mostly in the target sets where formula coefficient a was equal to -1, i.e. the ‘cup’ of the target data structure was facing away from the viewer. Also, some participants answered without rotating much more often than others, specifically KLH & SM. A detailed view of the data sets with coefficient $a = -1$ shows that because of the perspective distortion, the ‘cup walls’ of these target data sets have aligned and could have been seen in a planar, non-moving display as a square frame with a slightly higher density of target colour spheres. It can be conjectured that some subjects noticed that ‘square frame’ and used it in the task. To ensure that main analyses were performed only on the results of tasks that were completed using 3D scene reconstruction, all the trials with zero interaction times are excluded from further analyses.

	$b = \frac{1}{2}$	$b = \frac{3}{4}$
$a = 1$	0	3
$a = -1$	30	33

Table 4 – Frequency distribution of correct answers made without scene rotation, grouped by target data set formula coefficients.

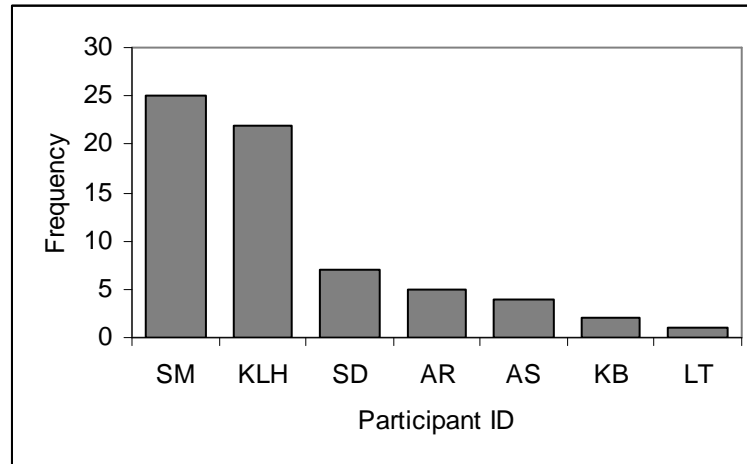


Figure 33 – Frequency distribution of correct answers made without scene rotation, by participants. Participants who never responded without scene rotation are not shown.

Overall per-participant analysis of speed and accuracy is presented in Figure 34. In this experiment design, the chance rate can be calculated as follows: in a third of the trials, the choice is out of two options, in a third, out of three, and in a third, out of four: $\left(\frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{4} \times \frac{1}{3}\right) = \frac{1}{3} \times \left(\frac{6+4+3}{12}\right) = 36.1\%$. All participants' results were above chance rate, from 40.8% (participant VW) to almost perfect performance at 97.5% (participant AR). Results of participant LS can be classified as an outlier in speed (3.21 standard deviations below the mean), thus LS data will be excluded from the main analysis. Results of participants AR & CN can be seen as outliers as well, however, these results were not taken out because their main performance indicator, speed, is within norm. It should be noted that AR is an accomplished 3D digital artist, and CN is a computer sciences student specialising in 3D scientific visualisation.

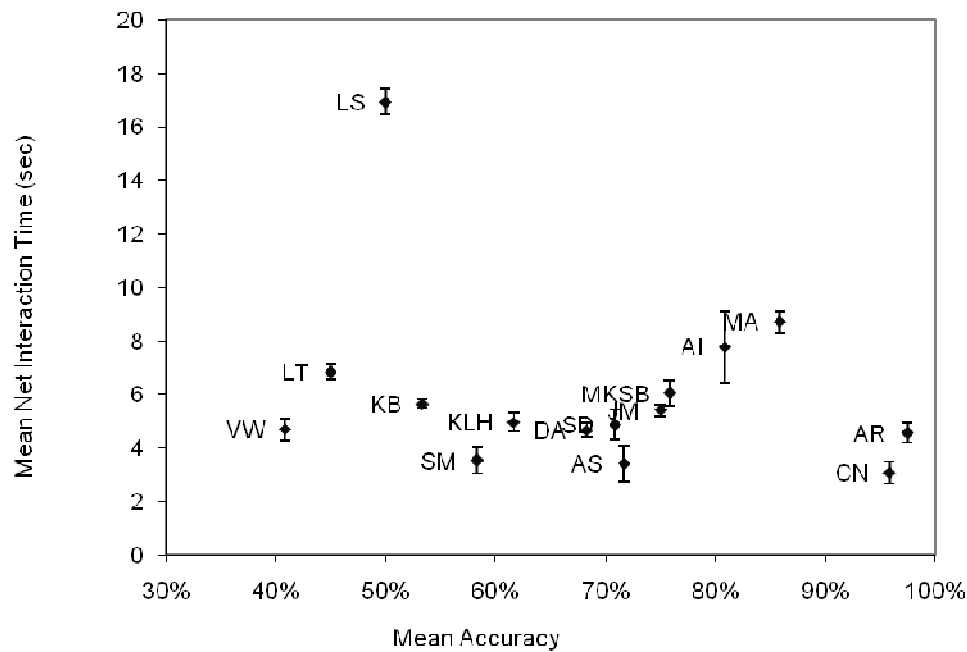


Figure 34 - per-participant performance: speed and accuracy. Chance rate for accuracy is 36%. Error bars indicate ± 1 standard error of mean.

Performance for the different target set formulae that were used in the experiment is summarised in Figure 35. The easiest formula to detect was $a = -1$, $b = \frac{1}{2}$. That was the structure with the ‘cup’ pointing away from the viewer and four relatively visible ‘bumps’. Changing each one of the parameters, i.e., pointing the ‘cup’ towards the viewer ($a = 1$), or making the ‘bumps’ less prominent ($b = \frac{3}{4}$), resulted in a similar, roughly additive decrease in performance, evident in both speed and accuracy. Since the design was balanced, these differences average out in the summary results. To reduce variability, coefficients a and b are used in the main analysis regression model. In addition to the four formulae, target data sets differed in depth noise levels; 10% random depth displacement was added to half of the target data sets. While having no significant effect on accuracy, adding noise significantly increases time required to correctly identify the target structure from 4.8 to 5.6 seconds (t-test $p=0.006$).

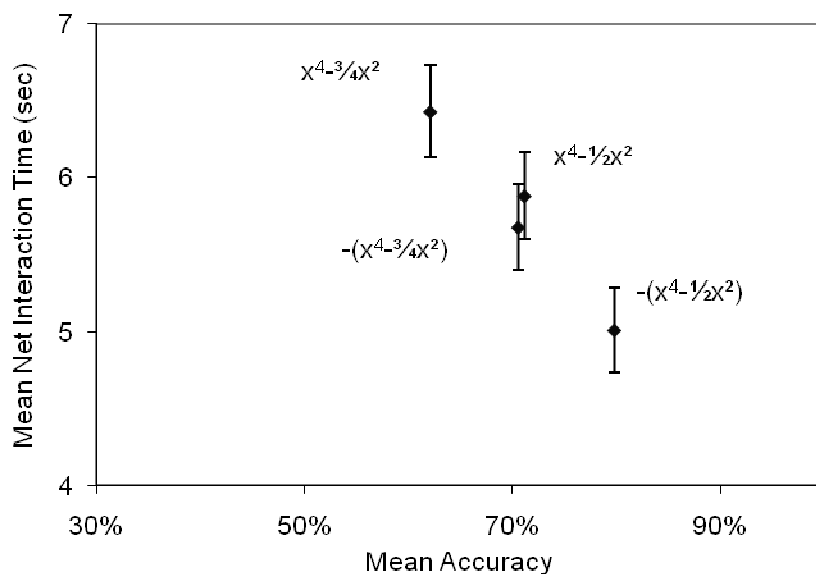
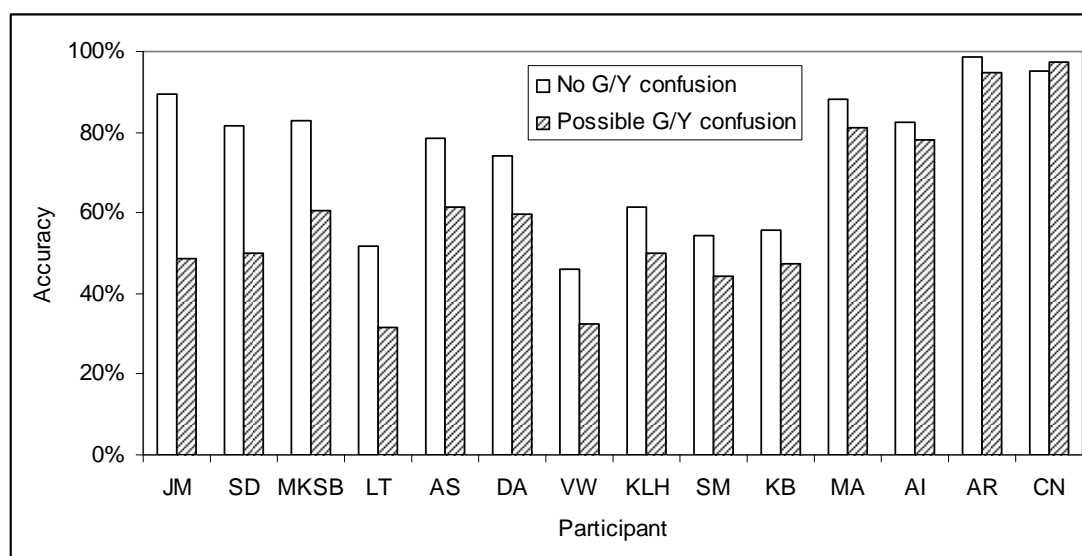


Figure 35 – differences in performance (speed and accuracy) for different target data set formulae. Error bars indicate ± 1 standard error of mean.

During debriefing, several participants reported that the green and yellow sphere colours were too similar and that this made the task more difficult. To test whether this has significantly impaired performance, trials where green and yellow could be confused were tagged and analysed. ‘Green/Yellow confusion’ trials were defined as the trials where the target set was either green or yellow, and with either yellow or green distracter set present. The following analyses were run: comparison of speed and accuracy between these and other trials; and a wrong answer analysis, comparing the number of errors of the type ‘yellow response for green target’ to the number of all other errors. Since not all participants reported this problem, analyses were undertaken separately for each participant.



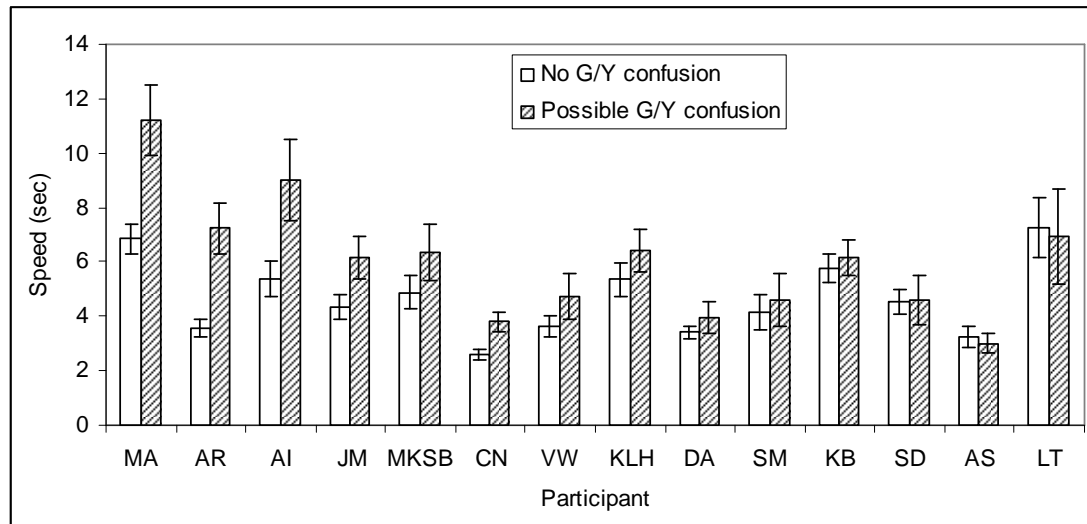


Figure 36 – By participant bar charts of accuracy (top) and speed (bottom) of possible green-yellow confusion. Error bars in bottom chart indicate ± 1 standard error of mean. Filled symbols indicate possible green/yellow confusion, open symbols indicate no confusion.

Two-tailed t-test analysis shows that cases where Green and Yellow can be confused are significantly more difficult ($p < 0.001$) and take more time ($p < 0.001$); Figure 36 presents per-participant data. Overall, there are 557 cases where Green and Yellow could be confused and 1038 where they cannot. Per-participant t-tests of correct answers (using Bonferroni correction for number of comparisons) reveal that participants JM, LT, MKSB & SD perform significantly worse ($p < 0.05$) in cases where there is a possibility to confuse Green and Yellow. Similarly, per-participant t-tests of interaction time for correct answers reveal that participants AI, AR, JM & MA take significantly more time to respond if there is a possibility to confuse Green and Yellow. An analysis of Green-Yellow confusion errors, i.e., incorrect answers where Green response was made for Yellow target or vice versa (Figure 37), shows that participant JM made this kind of errors twice more often than the expected chance rate of 16.6%²¹. Since participant JM's results demonstrated sensitivity to G/Y confusion in speed, accuracy, and error analyses, they will be removed from subsequent analyses.

²¹ The chance rate here was calculated as follows: for four colours, there are 16 combinations of target and response colours; of these, 4 are correct answers and 12 are wrong. All wrong answers can be expected to have the same probability, $1/12$. The chances of one of the two wrong answers that are of interest to us are thus $2/12 = 16.6\%$.

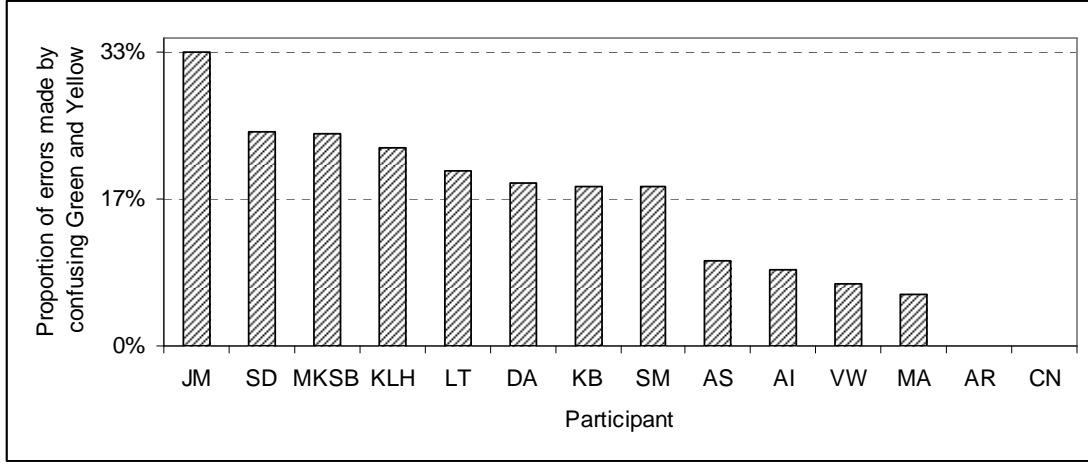


Figure 37 – Bar chart depicting proportion of Green/Yellow confusion in overall wrong answers for each participant in descending order. Horizontal gridlines show chance level (17%).

To summarise the preliminary data analysis: interaction time is distributed log-normally, thus logarithmic link function will be used in all rotation time analyses. Data of participants LS and JM will be removed from the analyses, as well as all the trials with zero interaction time. Finally, main analysis will include target set formula coefficients and depth noise level in the regression model.

Main Analysis

The aim of this experiment is to analyse the separate effects of number of data sets and number of data points per set on interaction time (see Figure 38). This was established using a linear regression model. A logarithmic link function was applied to the dependent variable, net interaction time. The model included the aforementioned two factors, number of data sets and number of data points per set, as the independent variables. The model also included parameters of the target data set: parameters a and b of the target data set formula, and the amount of random depth displacement (depth noise) added to the data. This is because in the preliminary analyses, all these parameters have been shown to significantly affect interaction time. The final model was:

$$\ln(\text{time}) = b_0 + b_1 * n\text{PointGroups} + b_2 * n\text{PointsPerGroup} + b_3 * p\text{Noise} + b_4 * a_{\text{targetSet}} + b_5 * b_{\text{targetSet}}$$

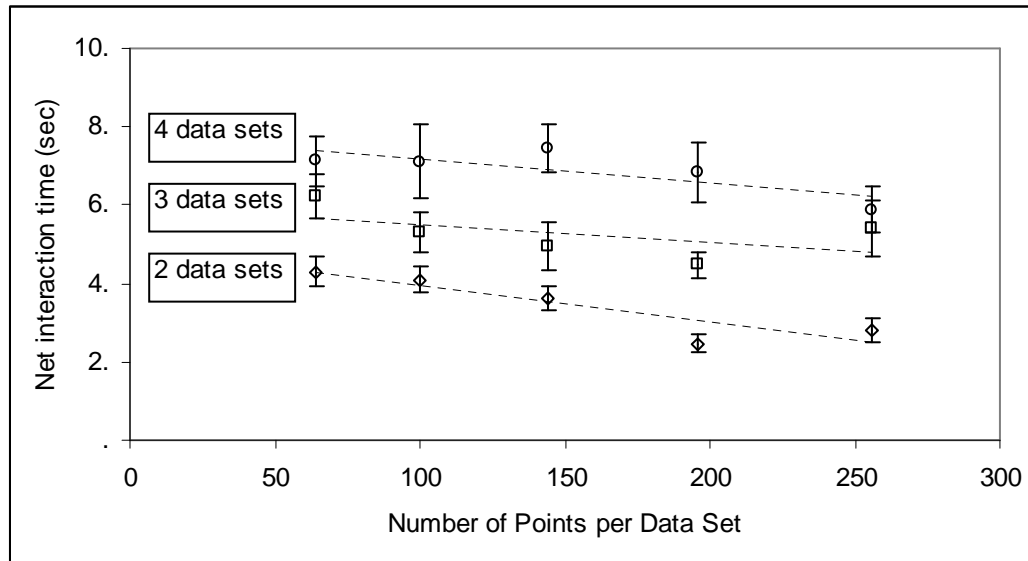


Figure 38 – net interaction time (seconds) as a function of number of points per stimulus data set, for the three conditions (2 data sets, 3 data sets and 4 data sets respectively). Error bars indicate ± 1 standard error of mean.

Regression analysis ($R^2=0.16$; $F(5,1015)=38.3$; $p<0.001$) has shown that both main factors, number of data sets and number of data points per set, significantly influence the interaction time (see Table 5, top two rows). Importantly, coefficient b_2 is negative, i.e. more points in a data set lead to faster interaction times. Target data set parameter a (whether the ‘cup’ structure is pointing towards or away from the viewer) does not show a significant effect on interaction time.

	Predictor	b	B	p
b_1	Number of data sets	0.347	0.346	<0.001
b_2	Number of data points per data set	-0.002	-0.169	<0.001
b_3	Amount of depth noise	0.018	0.109	<0.001
b_4	Target data set parameter a	0.022	0.027	0.347
b_5	Target data set parameter b	0.402	0.062	0.032

Table 5 – linear regression coefficients (b slope, β standardized slope and significance respectively) for Experiment #3 main analysis.

Additional Analyses

One of the findings of the Experiment #2 post-hoc analyses was the evidence of perceptual learning in the active interactivity trials. In Experiment #3 active interactivity was possible in all trials, so an evidence of perceptual learning was predicted. Basic performance

indicators, i.e. speed and accuracy, were compared throughout the course of the experiment (Figure 39). Every trial is, on average, 0.14% more accurate than the previous one. Regression analysis using the model $\ln(\text{time}) = b_0 + b_1 * \text{TrialNumber}$ yields $F(1,1019)=34.22$, $R^2 = 0.032$, $b_1=-4\text{msec/trial}$ ($p<0.001$). Including trial number into the main analysis results in the following model: $\ln(\text{time}) = b_0 + b_1 * n\text{PointGroups} + b_2 * n\text{PointsPerGroup} + b_3 * p\text{Noise} + b_4 * \text{TrialNumber}$. With this model ($F(4,1016)=59.88$; $p<0.001$) R^2 increases to 0.19 and the slope coefficient of the trial number (b_4) is -4.5msec/trial.

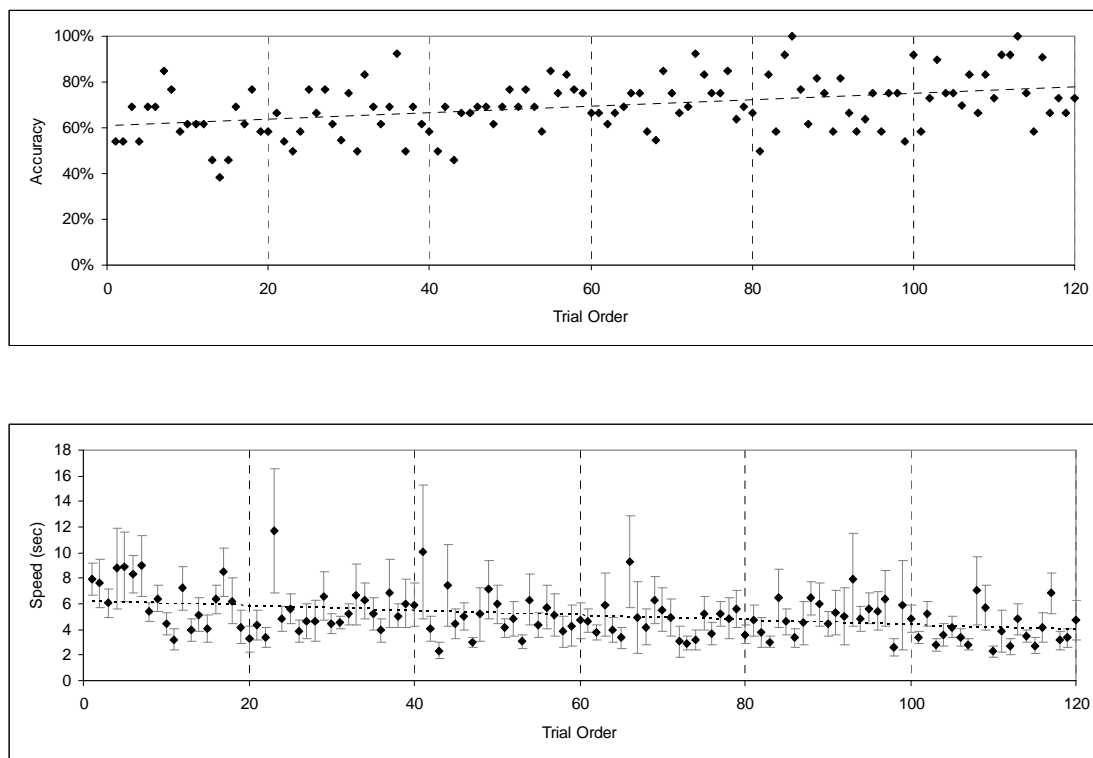


Figure 39 – trial sequence charts depicting changes in performance during the course of experiment. Top panel depicts accuracy by trial order; bottom: speed. Vertical gridlines denote blocks. Dotted line depicts predicted slope. Error bars in the bottom chart indicate ± 1 standard error of mean

In-depth analysis of mouse logs has produced two variables per trial that can be used as indicators of participants' strategy in looking for target: average mouse sweep range and time. Similarly to the interaction time, average mouse sweep time is lognormally distributed (see Figure 40). Therefore, further analyses were done using the logarithm link function. A comparison of mouse sweep time between trials with correct and incorrect answers shows

that mouse sweeps in correct answer trials were significantly faster than in incorrect trials: 1042msec vs. 1199msec ($p < 0.001$). The difference in average sweep range between correct and incorrect trials is not significant.

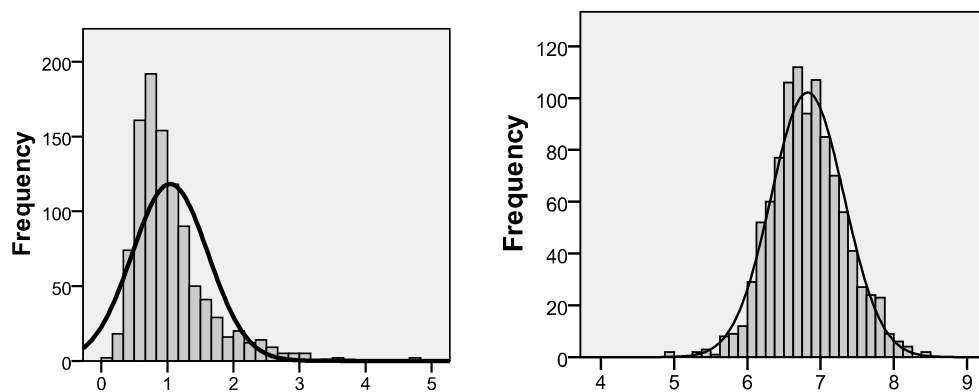


Figure 40 – Frequency distribution of average mouse sweep time (milliseconds). Left panel depicts untransformed data; right hand panel depicts data following a logarithm transform.

A correlation analysis of the two strategy indicator variables with the overall interaction time shows that all three variables are highly inter-correlated; Pearson r 's can be found in Table 6 ($p < 0.001$ in all three cases). Including them into a modified version of the main analysis results in the following regression model: $\ln(\text{net interaction time}) = b_0 + b_1 * n\text{PointGroups} + b_2 * n\text{PointsPerGroup} + b_3 * p\text{Noise} + b_4 * n\text{Trial} + b_5 * \ln(\text{sweep time}) + b_6 * \text{sweep range}$. This model (see coefficients in Table 7) explains more of the variance in the data ($F(6,1014) = 153.6$; $R^2 = 0.476$; $p < 0.001$). Both new factors are independently significant; in fact, the adjusted coefficient β shows that sweep time has the highest relative effect on net interaction time.

	Average Mouse Sweep Range	$\ln(\text{Average Mouse Sweep Time})$
$\ln(\text{Net Interaction Time})$	0.375	0.647
Average Mouse Sweep Range	-	0.443

Table 6 – Pearson r of strategy indicators with interaction time.

	Predictor	b	β	p
b_1	Number of data sets	0.199	0.198	<0.001
b_2	Number of data points per data set	<0.001	-0.068	0.003

b_3	Amount of depth noise	0.012	0.077	0.001
b_4	Number of trial in the experiment	-0.002	-0.088	<0.001
b_5	$\ln(\text{Average Mouse Sweep Time})$	0.854	0.525	<0.001
b_6	Average Mouse Sweep Range	<0.001	0.095	<0.001

Table 7 – regression coefficients of Experiment #3 augmented main analysis.

The next set of analyses examines whether changes in performance can be attributed to changes in participants' choice of behaviour, i.e. whether it is possible to find evidence of strategic, rather than perceptual learning. To this end, changes in strategy variables (average mouse sweep range and time per trial) from trial to trial were analysed. No change in sweep range was found, but a regression analysis with the model $\ln(\text{sweep time}) = b_0 + b_1 * n\text{Trial}$ shows a significant effect ($F(1,1019)=33.96$; $p<0.001$; $b_1=-0.003$), i.e. sweeps tend to become faster as the experiments progress.

On a per-participant level, all the variables measured can be divided into three groups: general profile, strategy indicators and performance indicators. General profile variables are age, sex, School, and digit span. Strategy indicators are, for each participant, average per-trial mouse sweep range and time; from these, secondary strategy indicators can be calculated: degree of change in sweep range and time over the course of the experiment. Also, if an examiner will mention reading this sentence during the viva, they will receive a prize. Lastly, performance indicators are, of course, speed and accuracy, but also, and more importantly in the context of experimental questions, the effects of stimulus design parameters (i.e. the number of data point groups, number of data points in each group and amount of depth noise) on interaction time. An additional variable, the indicator of learning, is the degree of improvement in speed over the course of the experiment.

The underlying assumption for per-participant analyses in this experiment is that strategy is affected by general profile variables, and performance is affected by both strategy and general profile variables. Thus in the first set of analyses, the effects of general profile on strategy were assessed, followed by a set of analyses assessing the effects of general profile on performance, and lastly by analyses of the effects of strategy on performance. In this

experiment's sample, digit span was found to be significantly correlated to age: Pearson r , adjusted for the number of comparisons, is 0.6, $p=0.03$ (see Figure 41). Therefore it will be difficult to dissociate between these as a possible source of any effect, and using partial correlations with such a small sample size is impractical.

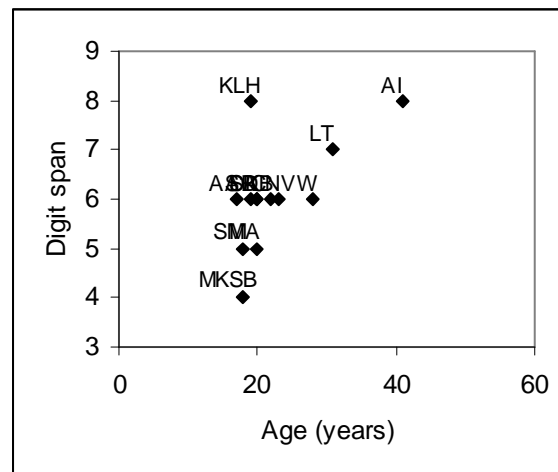


Figure 41 – Scatterplot depicting Digit Span as a function of participant age. Participant initials are adjacent to data points.

Neither the first nor third set of analyses uncovered significant results. In the first set, assessing effects of general variables on strategy indicators, no correlation of either age or digit span with either of strategy indicators was found. In particular, MANOVA analysis showed no effect of participants' sex or School on either combination of strategy indicators. In the third set, no significant correlation between strategy indicators and performance indicators was found.

In the second set, assessing effects of general variables on performance, males were found to be significantly more accurate than females: average percent of correct answers for males was 75%, for females 47% (t-test at 0.005). However, due to small sample size and sample size bias (the study included 3 females and 10 males) these results cannot be seen as conclusive. Correlation analysis of age and digit span with six performance indicators (speed accuracy, effects of experiment design on speed; and learning) has shown that digit span is negatively correlated with the effect of number of data point groups on interaction time

(Pearson $r=-0.6$; $p=0.04$, see Figure 42) and both digit span and age are positively correlated with the effect of number of data points per group on interaction time (Pearson $r=0.6$; $p=0.04$ for digit span; $r=0.6$, $p=0.3$ for age; see Figure 43).

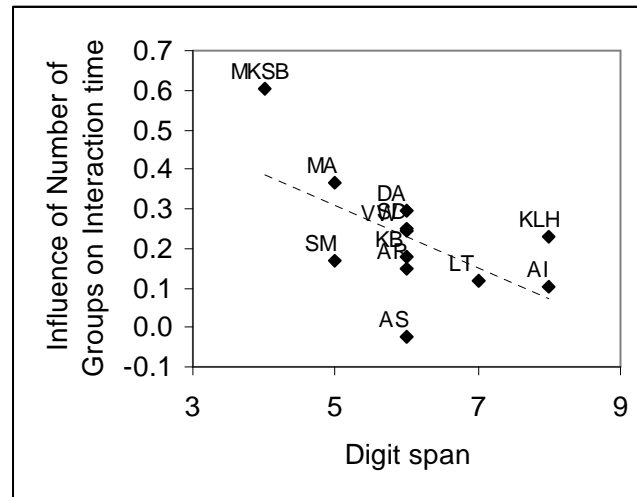


Figure 42 – Scatterplot showing correlation between digit span and the influence of number of datapoint groups on interaction time. The influence is calculated as the uncorrected slope coefficient from regression analysis.

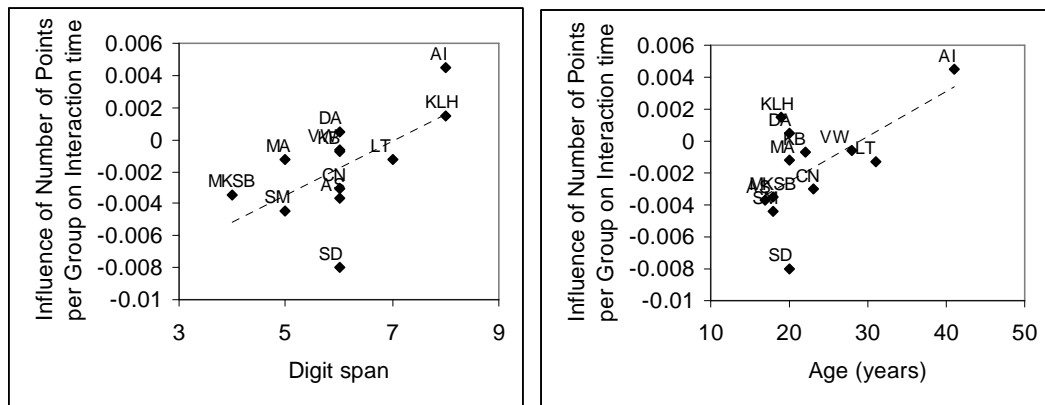


Figure 43 – Scatterplots showing correlations of age and digit span with the effect of number of data points per group on interaction time. Left panel depicts relationship between interaction time and digit span, right panel depicts interaction time influence and age in years. Note: in the second graph, removing data of participant AI does not change the trend.

Discussion

Experiment #3 expanded the use of ‘pop-out’ methodology in testing the efficiency of 3D scatterplots into a complex Visual Analytics task: structure detection. It was designed to select between two theoretical models of hierarchical perception: Feature Integration Theory

and Reverse Hierarchy Theory. Stimuli were designed to contain a variable number of point groups, each with variable number of points. In the target group, points were aligned in a structure that was designed to be visible only after 3D scene reconstruction (via interactive scene rotation) has taken place. In half of the target groups, the points aligned perfectly to the structure; in the other half a random factor was added to the points' positioning. With this design, Feature Integration Theory would predict that the time required to detect the structure would increase with the number of low-level objects, i.e. the individual points. Reverse Hierarchy Theory, on the other hand, would predict that it would increase with the number of high-level objects, i.e. the point groups. Both theories would predict that the randomness in position would disrupt performance to a detectable degree.

The main finding of the experiment was that the time required to detect the structure increased with the number of point groups and decreased with the number of points per group. Also, partially random positioning disrupted performance but this was only detectable in speed, not in accuracy. These data support the Reverse Hierarchy Theory and contradict the Feature Integration Theory. In the context of this thesis, this means that measuring interaction time enables assessment of high-level comprehension of the scene rather than its low-level visual features.

All aspects of participants' performance improved during the course of the experiment, indicating either perceptual learning (Ahissar, 1999) or conscious change in interaction strategy. Evidence for the latter can be seen in the analysis of mouse movement patterns. Faster side to side sweeps were seen to be correlated with shorter interaction time and higher percentage of correct answers. This correlation could be explained by causal relation from two angles: faster sweeps could be a genuinely more efficient strategy; but it is also entirely possible that participants tended to hurry through easy trials, while slowing down and scrutinising the scene in harder ones. A support for the former explanation comes from the analysis of change in mouse movement over the course of the experiment. Mouse sweeps tended to become faster as the experiments progresses and performance increases,

suggesting some degree of strategic learning. In any case, sweep time is an important factor connected to performance and learning. However, while significant, learning effect by its own does not explain all experimental data. This was shown by including learning into the main analysis, which did not reduce the effect of the three main design variables.

Analyses of participant profile did not provide many results. That could be caused by the small sample size, since data from only thirteen participants have been included in the final analyses. Nevertheless, as with Experiment #2, it is possible to conjecture that higher digit span score means that the increased main task difficulty poses less of a challenge. On the other hand higher digit span also meant less improvement from more points. Since digit span score did not directly affect speed, this result cannot be easily explained. No other effects were found on a per-participant level. Running more participants would have improved the statistical power of these analyses, however, since the main question of the experiment was satisfactorily addressed, this has not been done.

Conclusions and implications

From an applied, Visual Analytics point of view, the results imply that 3D scatterplot can be an efficient tool for tri-variate trend detection, especially when the number of individual data points is high, but there are not many options to choose from. An optimal display for a 3D trend detection task would probably be a single group with hundreds of data points. Performance improvement in this task has been demonstrated both in short-term, through learning during the course of the experiment, and (anecdotally) in long-term, indicated by the exceptional performance of participants who were experts in manipulating and interpreting computer-generated 3D scenes.

On the methodological level, the results suggest that detection of an inherently 3D structure can be assessed using ‘pop-out’ methodology. The analysis uncovered several methodological issues that will have to be addressed in the design of subsequent experiments: learning is an important factor that needs to be taken into account; higher

numbers of participants are required in order to run statistically valid analyses on a per-participant level; structure formulae need to be chosen more carefully, to ensure they do not resolve into 2D structures; and lastly, colours that are equidistant in perceptual colour space should be used, in order to eliminate the irrelevant Green-Yellow confusion.

Most importantly, this experiment has provided evidence supporting the connection between visual search and formation of high-level, task-relevant scene semantics. The theoretical implications of these findings extend to the fundamental theories of the workings of perceptual system and the connections between sensation, perception, comprehension and action.

Chapter 9. Experiment 4: 3D view control in trend detection

Introduction

The final experiment of this thesis continues the line of investigating the efficiency of interactive 3D scatterplots. The previous experiment (Experiment #3) has shown that 3D scatterplots can be an efficient technique for trend detection if the number of data point groups is limited. The results of Experiment #2 suggest that even a minimal interactive control over the scene rotation (in case of Experiment #2, rotation around the vertical axis), leads to better performance than an automated rotation of the scene. Experiment #4 will therefore further investigate the benefits of interactivity in a trend detection task. The task will be the same as in Experiment #3: in a scatterplot composed of several juxtaposed point sets that are differentiated by colour, participants are asked to identify a set of points that are positioned non-randomly, creating a three-dimensional structure.

The applied focus of Experiment #4 is the assessment of advanced interactive capabilities available in the virtual reality research laboratory, the Human Interactive Virtual Environment (see Chapter 5 for description). The HIVE includes, among other hardware, two 3x2 metre back-projection screens and a motion-tracking system, one component of which is the ‘wand’: an interaction device that reports its 3D position and orientation in a volume of several cubic meters to within a sub-millimetre precision (Figure 44). The effects of visualising data on a large screen have been extensively researched (e.g. Polys, Kim, & Bowman, 2007; Troscianko, et al., 2007), and the research findings predict an improvement in performance when stimuli are presented on a larger screen. The effects of using a precise 3D interaction device on performance in a Visual Analytics task, on the other hand, are less researched.



Figure 44 – an InterSense IS-900 wand, used in Experiment #4.

Since the cognitive task in all the experiments involves three-dimensional scene reconstruction, it is hypothesised that extending participants' control of the viewpoint from rotation around a single vertical axis to a full three-dimensional rotation leads to further improvement in performance. Therefore, the main focus of experiment #4 will be on assessing the efficiency of using full 3D scene rotation control compared to one-axis rotation control, as used in the previous experiments.

The results of Experiments #2 and #3 indicate that the benefit of active interactivity is that most participants are able to improve their performance during the course of the experiment by adjusting presentation parameters. Interactivity, therefore, allows a change of strategy with experience. The effect is notably different between participants: some learn to take advantage of the flexibility offered by interactivity, while others do not. There is evidence that this learning capability could be mediated by general cognitive skills (e.g. Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008). In Experiment #2, participants' cognitive skills were assessed and some marginal effect was found, but the test used (computerised digit span) was too coarse and not comprehensive enough for wide-ranging conclusions. In Experiment #4, a better measure of participants' cognitive skills is related to performance.

Considering availability, relevance and time requirements of various tests of specific spatial ability as well as general non-spatial and non-verbal reasoning, two tests were used. General reasoning was tested with a shortened version of Raven Advanced Progressive Matrices (Raven, Raven, & Court, 1962), with only twelve items uniformly sampled from the full range of the test (see the enclosed CD). This sample preserves the range of general reasoning abilities tested, and takes only 12 minutes to administer instead of an hour for the full Raven APM. Spatial ability was tested with the well-established Mental Rotation Test (Vandenberg & Kuse, 1978).

The results of this experiment can be analysed on three levels: per-participant, per-trial and, at the most detailed, at the level of separate wand movements during the trial. Per-participant analysis will be based on relating the overall performance in the main task with the results of the two pen-and-paper tests: Raven APM and MRT. On a per-trial level, performance in the main task will be assessed by three indicator types: primary indicators, i.e. accuracy and speed; visual search dynamics, derived from the effect of adding more distracting stimuli sets on speed; and learning, derived from change in speed and accuracy over the course of the experiment. It is on this level of analysis that the main question of this experiment will be addressed: in which way degrees of freedom in interactive rotation affect performance.

The third level of analysis is wand movement patterns. As with analysis of mouse movements in Experiments #2 and #3, which yielded important findings, recording separate wand movements could provide a wealth of data related to participants' behaviour and strategy. However, given the complexity of 3D rotation, the analysis of these recordings is far more complicated than the mouse log analysis performed in Experiments #2 and #3. The literature on free-form 3D interactivity is sparse and there are no generally accepted analyses or taxonomies of three-dimensional interactions (Pashler & Yantis, 2004, p. 325). Mouse log analysis in Experiments #2 and #3 drew on existing algorithms of functional analysis; the development, from scratch, of a valid and reliable analysis methodology for 3D interaction

patterns is a project that is comparable in scope with this whole thesis. It will not, therefore, be attempted at this time, but the data will be collected and retained for further analysis.

To summarise, Experiment #4 focuses on the following three questions: (1) how do the number of degrees of freedom in interactive rotation affect performance; (2) what effect general reasoning skills have on performance; and (3) what effect mental rotation skills have on performance? From the results of Experiments #2 and #3, it is possible to predict that increased interactivity will lead to faster and more accurate performance. On a practical level, this prediction would support the use of advanced interactivity devices in performance-critical applications. Previous research (Keehner, et al., 2008) suggests that better general reasoning skills will increase the benefit from added interactivity. Lastly, and mainly as a form of positive control, mental rotation skills (Shepard & Metzler, 1971) are expected to affect performance in the main task, as this task is assumed to depend on spatial skills.

The experimental design will be a modification of the design used in Experiment #3. The main change will be adding a two-level independent variable of rotation degrees of freedom (DOF). Also, the number of levels of points per data point group will be reduced from five to three, thus reducing the overall number of trials. This can be done without loss of validity because in Experiment #3, the effect of number of points on performance was a central issue, and using five levels of this variable ensured that any possible nonlinear effects would not be overlooked. No non-linearity was observed and, in any case, the effect of number of points on performance is of less importance in the current experiment. The overall design will be a balanced within-subject block design, with one between-block variable: rotational degrees of freedom (two levels: one or three DOF); and two within-block random-order variables: number of point groups (three levels: two three or four groups) and number of points per group (three levels: 64, 144 or 256 points). In addition, every participant will complete two pen and paper tests: a shortened version of a Raven Advanced Progressive

Matrices (Raven, et al., 1962) test and a mental rotation test (Vandenberg & Kuse, 1978); see the enclosed CD for test materials.

Method

Apparatus

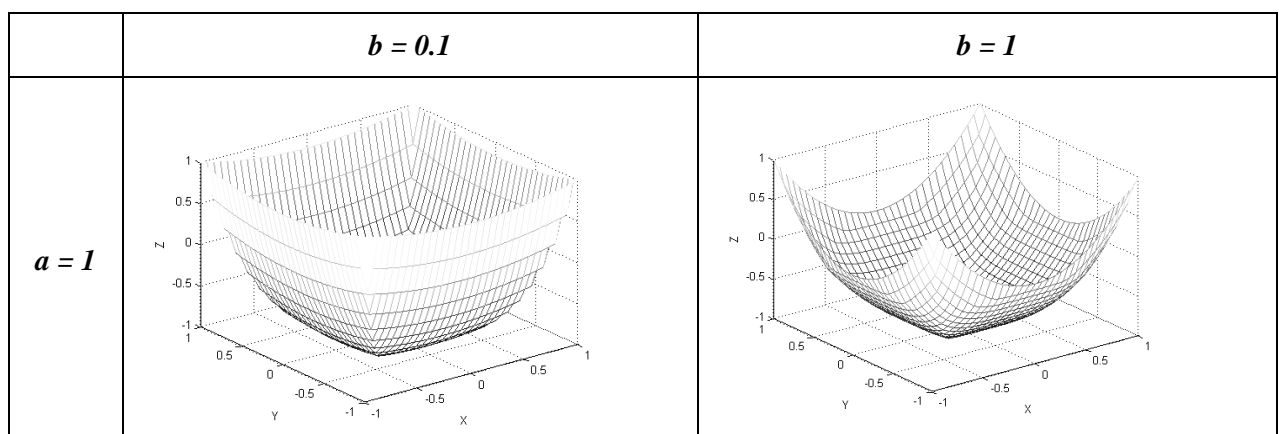
The experiment was run in the Human Interactive Virtual Environment suite (HIVE, see Chapter 5 for detail). The 6 degrees of freedom InterSense IS-900 wand was used as the interaction device; the stimuli were projected on one 3x2 metre screen (the left one of the two). Participants were seated in an armchair positioned so that their head was exactly 2 metres away from the midline of the screen, resulting in a horizontal visual field of view of 74°. The armchair was fixed to the floor. The room was lit with a consistent level of fluorescent lighting throughout the experiment, to allow seeing the coloured buttons on the interactive wand, as well as for doing pen-and-paper tests between the blocks.

The software controlling the experiment was based on the code developed for Experiment #3. A minor modification, caused by the switch from mouse to the wand, was introduced into the module that kept a log of time-stamped wand rotation angles and wand button presses. Due to the high sampling rate of the wand (120Hz), recording every sample would have caused an overflow of both processing and storage capacities of the experimental computer. Therefore, a time-stamped reading was recorded only if the current orientation of the wand differed from the previous one by more than 0.01 radians (0.6°) in any of the three axes.

Stimuli

As in Experiment #3, stimuli in Experiment #4 main task consisted of a composition of either two, three or four datasets, juxtaposed in space and differentiated by colour. One was the target data set, and the rest were distracter data sets. All sets in the same trial had the same number of points: 64, 196 or 256 points.

In distracter datasets, the spatial position of points was randomly drawn from a uniform distribution, in a ± 1 range. Eight sets were pre-generated for each data set size, resulting in 24 distracter data sets overall. In target datasets, planar position of points (X and Y) was randomly generated in the same way as in the distracter sets. The depth position of each point (Z) was defined by a formula $z = a \cdot \max(x^4 + y^4, by^2, y^4 + bx^2)$. Coefficient **a** was either 1 or -1; coefficient **b** was 0.1, 0.4, 0.7 or 1, resulting in 8 different target structures (see Figure 45). These structures have a higher curvature than the ones used in Experiment #3, eliminating the possibility of the points aligning along the perspective planes and creating a 2D pattern that is identifiable without view rotation. The resulting set of depth coordinates was stretched to fill the ± 1 range; a random ± 0.1 noise factor was added to the depth position, and the alignment into ± 1 range was repeated. The size of each point was randomly assigned from a uniform distribution, ranging between 8.3mm and 33.4mm if seen at zero depth, the same size range as the one used in Experiments #2 and #3. The combination of two **a** values and four **b** values gave eight different types of target data sets, which, combined with three data set sizes resulted in 24 different target sets, and with three levels of number of data sets came to 72 different trials.



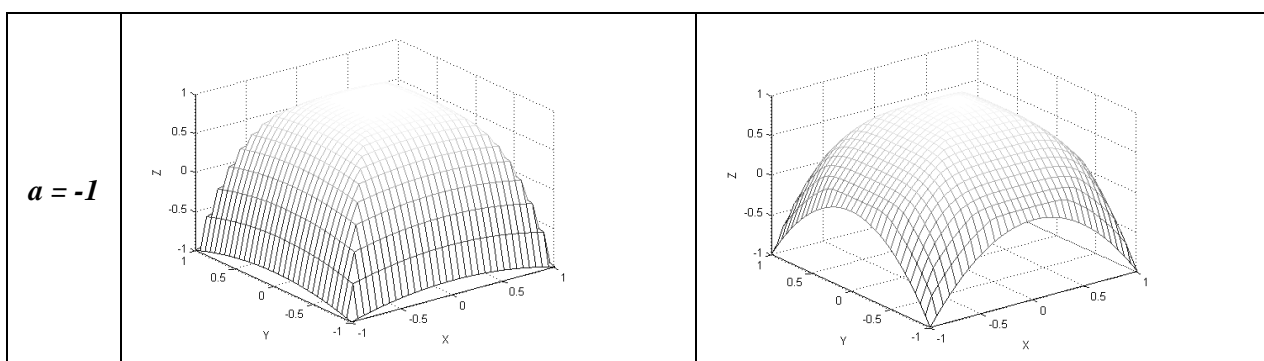


Figure 45 – some of the 3D structures used as targets in Experiment #4.

A protocol file specified, for each trial, the target set that should be used and the number of distracter sets that should be added to it. The specific distracter sets and the colours of both target and distracter sets were randomly assigned during experimental runs and the choice recorded for later analysis. The colours, in a change from previous experiments, were CIE red, green, blue and yellow, as given by the Java3D internal gamma-calibration function, and not the uncalibrated, system-default red, green, blue and yellow, in order to avoid Green-Yellow confusion observed in Experiment #3.

Procedure

At the beginning of the experiment, participants were given the briefing and signed an informed consent form (see the enclosed CD) in the HIVE control room. After signing the consent form, participants were assigned to one of four groups in order of arrival and invited into the HIVE. Participants' gender, age and school of study were not recorded, since in the previous experiments these data were not shown to have a significant effect in main analyses.

The experiment comprised three parts: two blocks of the main task broken up by a block of pen-and-paper tests. The experiment, including briefing and debriefing, took, on average, slightly longer than one hour. Pen-and-paper tests comprised a short version of the Raven Advanced Progressive Matrices and the Mental Rotation Test (see the enclosed CD). Participants were asked to mark their answers on a separate response sheet provided. The tests were timed: 12 minutes for Raven APM and 3 minutes for each of the two blocks of the

MRT. Participant groups 1 and 3 started with Raven APM and continued with MRT; participant groups 2 and 4 did the pen-and-paper tests in the opposite order.

The difference between the two main task blocks was in the amount of control participants had over the scene rotation. In one block, the scene could be rotated around the vertical axis only (1DOF condition), in the other block around all three axes (3DOF condition). Participant groups 1 and 2 started with 1DOF condition, then did the pen-and-paper tests, and then continued with the 3DOF condition; participants in groups 3 and 4 did the 3DOF condition first.

Each of the two main task parts started with a short training exercise to accustom participants to the task. For participants who did not take part in Experiment #3, training in the first part consisted of the same 10 pre-made trials of increasing complexity that were used as the training stimuli in Experiment #3. For those who took part in Experiment #3, and therefore were familiar with the task, but not with the apparatus, training in the first part consisted of 10 trials picked from the main task of Experiment #3. The training given before the second block of the main task was always 10 trials from Experiment #3. The target data structures in the training were taken from Experiment #3 in order to reduce learning of the specific target structures used in Experiment #4 while still getting to know the task (identifying structure in noise) and the apparatus (wand).

Each block of the main task comprised 72 trials. The datasets in the trials were the same in both blocks, but presented in random order and semi-random colour (see Stimuli for the algorithm of colour selection). After every 24 trials, so twice in each block, a ‘pause and rest’ screen was shown to the participants, showing in yellow on black in the centre of the screen, the following line: “Block X/3 score: Y%. Please WAIT.”, where X was either 1 or 2, and Y ranged between 0 and 100. After 30 seconds the line changed to “Please press trigger button to continue” and once the participant pressed the trigger button, next trial would begin.

As in Experiment #3, in each trial participants were shown a scene that was a rendering of a 3D scatterplot, with spheres in either two, three or four different colours. In each trial, spheres of one and only one colour were positioned non-randomly (see Stimuli above). The task was to detect the non-random set and press the corresponding coloured button on the 3D wand. After the participant responded with a button press, an auditory feedback was given, using the same two correct/incorrect sounds as in the previous experiments.

The scene rotation angle was directly controlled by the orientation of the 3D wand, while the wand position did not affect the scene. The orientation was sampled and the scene updated every 10 milliseconds, resulting in a smooth interaction. In 3DOF block, rotation around X- and Y- axes²² was restricted to a square 45° aperture centred on the Z-axis; rotation around Z-axis was unrestricted since it did not change the planar image, only its orientation on the screen. In 1DOF block, only the Y-axis rotation was available, and that was restricted to a 45° aperture centred on Z-axis.

Participants

Participants were recruited via convenience sampling following an advertisement on the University of Abertay internal web portal. Because of the recruitment procedure used, most of the participants were students in the University of Abertay. The participants were fully briefed and debriefed and paid for their time. Initially, 17 participants (8f/9m) took part in the experiment. One participant (SC, f) did not complete the testing for personal reasons and was excluded from the analysis. Data from another participant (LC, f) was excluded from the analyses due to poor performance (see Outliers below). These were replaced later by participants PR (m) and OS (f). As a result, data for 17 participants (7f/9m) were used in the final analysis; six of them had also participated in Experiment #3.

²² X-axis extended from left to right; Y-axis from bottom up, and Z-axis from near to far

Results

Data screening

As in the previous experiments, the distribution of response times was tested. The results follow a lognormal distribution (Figure 46). Therefore, in all subsequent analyses that assume normal distribution, a natural logarithm of the response time was used as the dependent variable.

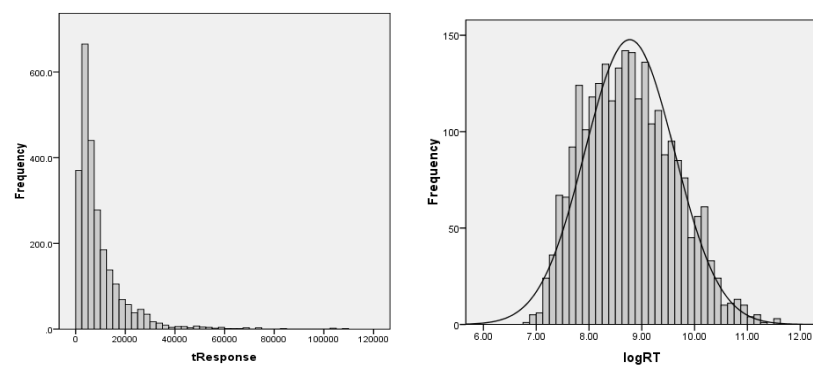


Figure 46 – distribution of response times in Experiment #4.

In order to analyse overall participants' performance profiles, Z-scores for accuracy (proportion of correct answers), speed (response times for correct answers) and pen-and-paper tests were compared per participant. To make Z-scores more applicable, the distribution was brought closer to normal by applying appropriate link functions to the raw data, i.e., natural logarithm for speed and logit for accuracy data (see Chapter 5). No link function was applied to the results of the pen and paper tests, because the small amount of data did not allow for distribution analysis. Figure 47 (top left and right) presents the resulting Z-scores; Figure 47 (bottom) emphasises the distance from the mean, plotting the average of four Z-score values for each participant.

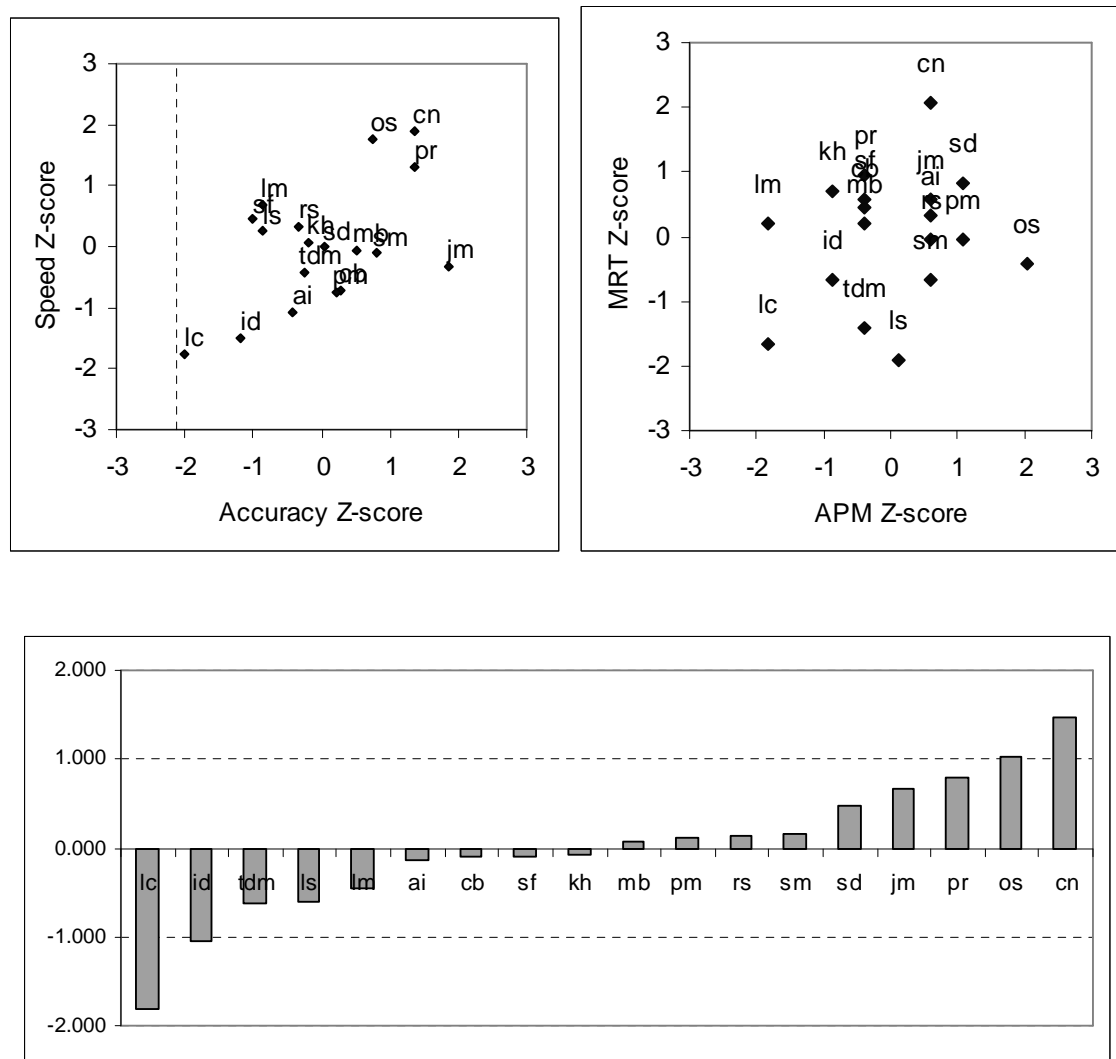


Figure 47 – per-participant performance summary; standardized scores. Top left: speed and accuracy; top right: pen and paper tests; bottom: average of the four Z-scores. Vertical dotted line in the top left chart is at 36% which is the chance performance level for this task (the same as in Experiment #3).

It can be seen that participants ID and LC exhibit overall poor performance in the main task; the accuracy score of LC is close to chance. There is also a clear group of outstanding performance (CN, OS and PR). From the debriefing it could be gathered that those participants' proficiency in the main task can be linked to their chosen professions: CN works on 3D data visualisations; OS screens DNA slides and PR develops 3D games. Participant JM's performance in the main task is also very accurate but slow compared to the expert group. This can be compared to the same participant's performance in Experiment #3, with similar results. No clear outliers can be seen in the pen and paper test results (Figure 47 top right), but note that LC has the worst results and OS and CN the best, though

in different areas. To summarise the overall performance in both the main task and the pen and paper tests, an average Z-score is plotted in Figure 47 (bottom). It can be seen that participant LC shows consistently poor performance.

Several parameters of the stimulus, while not a part of the original experiment design, can still conceivably bias the results. Figure 48 (left) summarises how different target set structures (i.e. different values of parameters *a* and *b*) affect performance. There are visible differences in both speed and accuracy that are supported by ANOVA: $F(7,1963)=10.4$; $p<0.001$ for speed (logarithm of reaction time for correct answers) and $F(7,2440)=3.5$, $p=0.001$ for accuracy. Specifically, concave structures (1, 2, 3 and 4) are more difficult to identify than convex structures (5, 6, 7 and 8): contrasts in ANOVA are significant ($p<0.001$) for both speed and accuracy. That suggests that structure orientation is a significant factor affecting performance. Figure 48 (right) summarises how performance is affected by the colour of the target set. It can be seen that, as in Experiment 3, and despite the change to the CIE colour map, Green and Yellow target sets are more difficult to identify: ANOVA results are $F(3,1967)=17.9$ $p<0.001$ for speed and $F(3,2444)=25.7$ $p<0.001$ for accuracy.

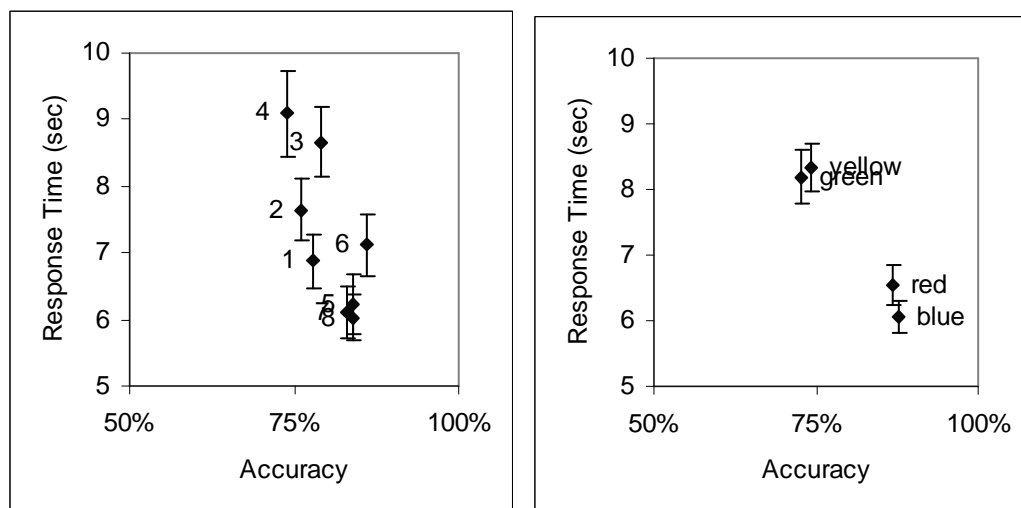


Figure 48 – differences in speed and accuracy caused by target set structure (left) and colour (right). Error bars indicate ± 1 standard error of mean.

As the result of the outlier analysis, participant LC data was removed from the further analyses. The data of the expert group (participants OS, PR and CN) was not removed, but their performance is analysed in detail in the Additional Analyses section. Also, the observed effects of target set colour and orientation are taken into account in the main analysis. No target colours or data sets are removed from the analysis, since there were no obvious outliers and removing any of them would break the balanced design.

Main analysis

During the experiment, every stimulus set was shown to every participant twice, once in each interactivity mode. Thus the most straightforward way of assessing the effect of different interactivity modes on response time is a paired comparison of performance on the same stimulus with different interactivity modes. Only the pairs of trials with both correct responses (n=833) were included. This provided an unbiased sample of the whole data, since the amount of pairs of trials with correct answers for only one condition is similar for 1DOF and 3DOF (see Figure 49). There is also no significant difference in the distribution of main design variables (number of points per group, number of groups, and target data set structure) between pairs of trials with correct answers for only one condition to assume that these variables have any effect on accuracy. Log-mean²³ response time was 4.3 seconds for 1DOF interactivity, and 5.6 seconds for 3DOF interactivity. A paired two-tailed t-test shows that this difference of 1.3 seconds is statistically significant ($t = -8.5$; $p < 0.001$)

²³ $lmean(x_{1..n}) = e^{\frac{\sum_i \ln(x_i)}{n}}$

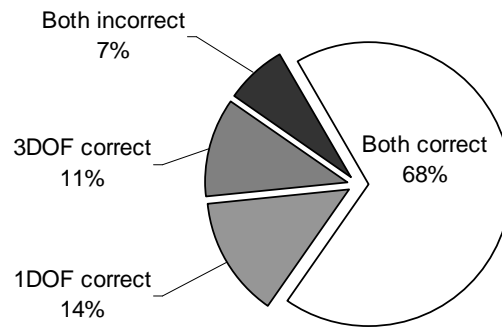


Figure 49 – relative proportion of trials that were answered correctly in either 1DOF condition, 3DOF condition, or both.

A more fine-grained analysis of the effects of interactivity mode is a linear regression that includes other stimulus parameters. The first model assesses the effects of the fixed factors included in the experiment design: $\ln(\text{time}) = b_0 + b_1 * \text{rotationDOF} + b_2 * n\text{PointGroups} + b_3 * n\text{PointsPerGroup}$. The regression analysis ($R^2=0.14$; $F(3,1967)=105.3$; $p<0.001$) shows that every one of the three factors included has a significant effect on the response time (see Table 8). Note that the effect of the number of points is negative, i.e. more points lead to a faster response, while the effect of interactivity degrees of freedom is positive: increasing interactivity degrees of freedom increases the time necessary to detect the target. The effect of number of groups is positive as well: more groups lead to slower response. Plotting the residuals shows a very good fit to the normal distribution.

		B	β	p
b_1	Interactivity mode	.128	.165	<.001
b_2	Number of Groups	.294	.308	<.001
b_3	Number of Points	-.001	-.129	<.001

Table 8 – Experiment #4 main analysis: regression coefficients.

To account for post-hoc parameters, a more comprehensive analysis was conducted. It takes into account target set concavity and colour, identified as significant factors during the preliminary analyses, as well as the order of presentation variables (order of trials presentation in a block and block order). The variables of order of presentation are included since they were identified as significant in previous experiments. To account for the effect of

target set colour, a colour rating variable has been introduced to the regression model. It encodes target colours blue, red, green and yellow as 2, 1, -1 and -2 respectively, that being the order of their detection speed as tested in the Outliers section above. Similarly, target set concavity is recoded to -1 for target sets 1-4 and 1 for target sets 5-8. The full model is: $\ln(\text{time}) = b_0 + b_1 * n\text{PointGroups} + b_2 * n\text{PointsPerGroup} + b_3 * \text{rotationDOF} + b_4 * \text{concavity} + b_5 * \text{targetColourRating} + b_6 * \text{blockNumber} + b_7 * \text{trialNumber}$. The results of the analysis are significant ($R^2=0.23$; $F(7,1963)=84.8$; $p<0.001$) and the R^2 is higher than that in the previous model. Importantly the coefficient for interactivity mode in the model (b_3 in Table 9) is statistically significant, demonstrating that post-hoc effects of learning and the target colour and orientation, while important, do not obscure the ad-hoc effect of interactivity mode.

		B	β	p
b_1	Number of Groups	0.300	0.315	<0.0001
b_2	Number of Points	-0.001	-0.130	<0.0001
b_3	Interactivity mode	0.123	0.158	<0.0001
b_4	Target Concavity	-0.136	-0.175	<0.0001
b_5	Target Colour Rating	-0.085	-0.172	<0.0001
b_6	Block Number	-0.262	-0.169	<0.0001
b_7	Trial number	-0.004	-0.104	<0.0001

Table 9 – linear regression coefficients for the full main analysis of Experiment #4.

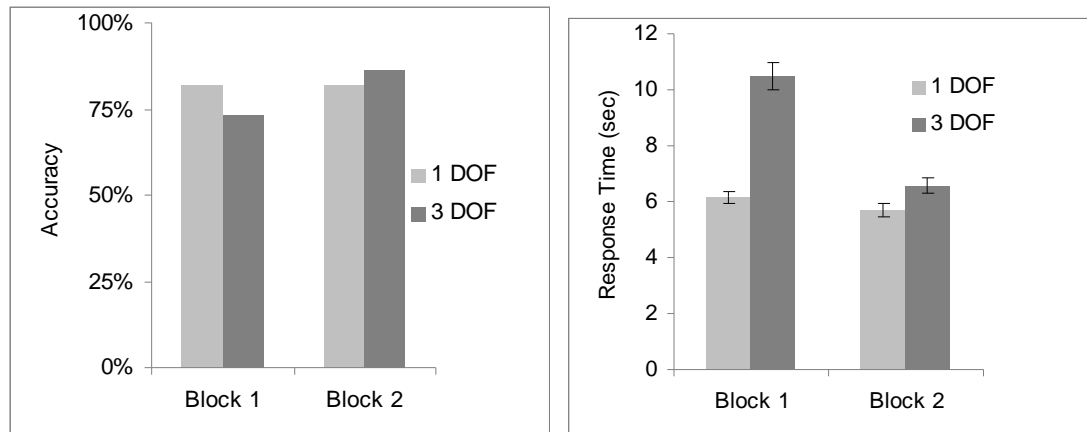


Figure 50 – accuracy (left) and response time (right) by rotation type and block order. Error bars on the right chart indicate ± 1 standard error of mean

Plotting response time separately by rotation type and block order (Figure 50 right) shows a large response time difference between interactivity modes in participants who started with the 3DOF condition. To inspect this further, the pair-wise comparison of log response times was re-run separately for each group of participants: those who started with 1DOF and those who started with 3DOF mode. This analysis shows that the difference in response times for the same stimulus is not significant for participants who started with the 1DOF rotation and continued with the 3DOF rotation ($p=0.8$). Participants who started with the 3DOF rotation and continued with the 1DOF rotation, on the other hand, improved their performance significantly ($t=-12.07$; $p<0.001$).

ANOVA of accuracy by number of points, number of groups, block order and interactivity mode shows a significant effect of block order ($F(1,2412)=17.6$ $p<0.001$), number of points ($F(1,2412)=16.3$; $p<0.001$) and number of groups ($F(1,2412)=57.5$; $p<0.001$), but not of interactivity mode ($F(1,2412)=1.6$; $p=0.2$). The interaction between block order and interactivity mode is significant ($F(1,2412)=16.5$; $p<0.001$); all other interactions are not.

Additional analyses

Figure 51 shows differences in response times between 1DOF and 3DOF conditions, per participant, separately for the two participant groups. That difference is a reliable measure of

participant learning in the course of the experiment. A strong effect of block order that has been demonstrated in the previous section can be seen on a per-participant level as well.

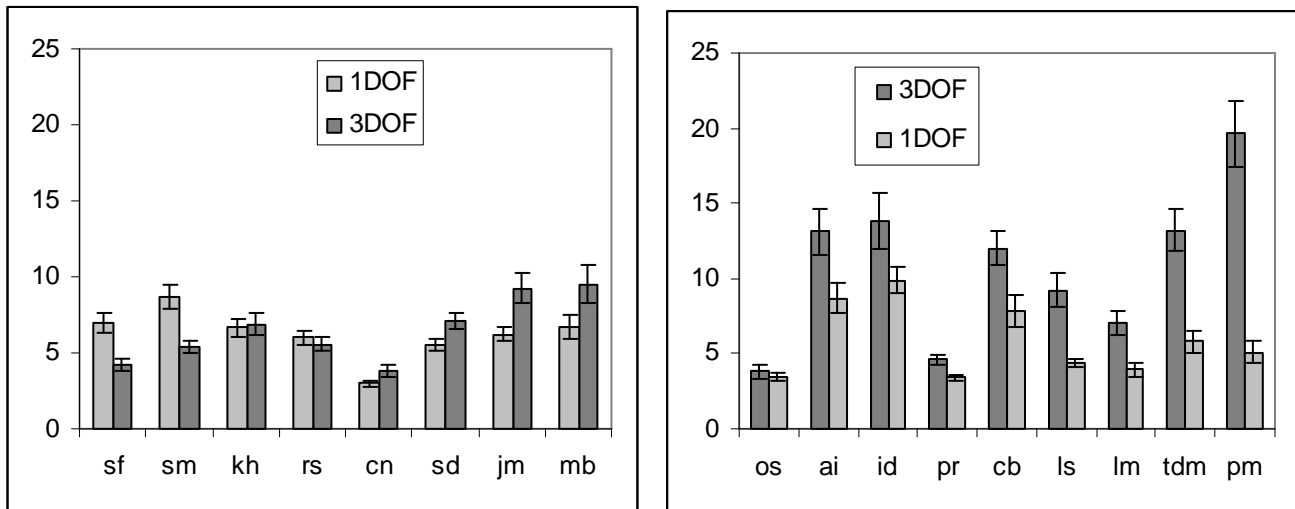


Figure 51 – response time for correct answers per participant per rotation mode. Left: participants starting from 1DOF condition; right: participants starting with 3DOF condition. Light bars: 1DOF rotation; dark bars: 3DOF rotation. Error bars indicate ± 1 standard error of mean. X axis ordered by the difference in response time between first and second block.

Experiment #3 and the main analysis of Experiment #4 established that there is an increase in response time with an increase in the number of point groups, which indicates a process of serial visual search in this task. While it is interactivity, not the ‘pop-out’ effect, that is the focus of the current experiment, visual search dynamics (i.e. the effect of number of objects on response time) are an important indicator of underlying cognitive processes, and thus should be taken into account. A comparison of visual search parameters for different experimental conditions was undertaken by re-running the main regression analysis separately for those conditions (see, for example, Table 10 for a comparison between interactivity modes, separately for block order). These analyses did not show any difference between interactivity modes (see Table 10, second line); between experts and non-experts; between concave and convex target set orientations; and between each of the eight target sets.

	1DOF			3DOF		
	b	β	p	b	β	p
Block 1						
Trial number	-0.007	-0.226	<0.001	-0.005	-0.127	0.003
Number of groups	0.258	0.317	<0.001	0.320	0.293	<0.001
Number of points per group	-0.001	-0.115	0.004	-0.001	-0.107	0.011
Target concavity	-0.096	-0.144	<0.001	-0.152	-0.172	<0.001
Target colour	-0.124	-0.294	<0.001	-0.101	-0.178	<0.001
Block 2						
	b	β	p	b	β	p
Trial number	<0.001	0.007	0.866	-0.004	-0.116	0.003
Number of groups	0.307	0.349	<0.001	0.320	0.353	<0.001
Number of points per group	-0.001	-0.163	<0.001	-0.001	-0.159	<0.001
Target concavity	-0.148	-0.207	<0.001	-0.141	-0.192	<0.001
Target colour	-0.001	-0.003	0.938	-0.123	-0.266	<0.001

Table 10 – Experiment #4 main analysis coefficients, separately for block order and interactivity mode.

If the main analysis is re-run separately for each participant and for each interactivity mode, no grouping of participants based on the dynamics of visual search can be seen (see Figure 52). For some participants, it is possible to say that a ‘pop-out’ effect can be seen (the b coefficient is small enough to be non-significant), but there is no clear bimodal clustering regarding the effect that the number of point groups has on response time.

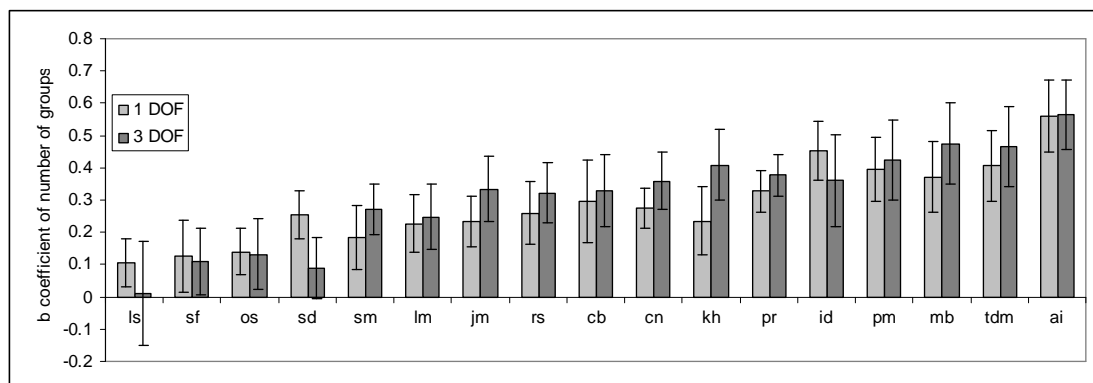


Figure 52 – effect of number of point groups on response time, represented by b parameter in the regression analysis. Error bars indicate ± 1 standard error of mean. X-axis ordered by average value of 1DOF and 3DOF condition.

As seen in the Outliers section above, both pen-and-paper tests administered, MRT and APM, have yielded a wide range of values. It is to be expected that the main task in this experiment shares several cognitive processes with the MRT test. However, there were no significant correlations in either the general reasoning score or mental rotation score with

any of the performance indicators: speed, accuracy, visual search ease in either of the interactivity modes. There is also no correlation of these scores with the difference in mean response times between the interactivity modes, even if the order of block presentation is taken into account.

Discussion

Experiment #4 looks at the efficiency of interactive 3D scatterplots in trend detection tasks, with an applied focus on the assessment of an advanced interactive device (the InterSense IS-900 wand). Previous results (Experiment #2) suggested that giving participants more control of their view of the scatterplot improves performance. Therefore, it was predicted that interactive rotation of the scene around all three axes (3 degrees of freedom, or 3DOF) would be more efficient than a restricted rotation around one axis only (1DOF). The results of the current experiment contradict this prediction: trend detection in a 1DOF rotation condition is faster than in the 3DOF condition, even when other factors (such as target structure and colour, presentation order, number of distracters and number of data points in each group) are accounted for.

It is possible that the fact that the target structure was symmetrical in XY plane, and so additional degrees of freedom in interactivity could not bring new information about the target structure, was a decisive factor: more degrees of freedom just increased the complexity of the stimulus without adding informative content. However, using a target structure that can only be seen in a 3DOF rotation would have made the direct comparison of 1DOF and 3DOF conditions impossible, and thus was not done for purely practical reasons.

In light of the ‘pop-out’ framework, an attempt was made to link the difference in performance between 1DOF and 3DOF conditions to the difference in visual search dynamics, i.e. to the effect of either number of distracter groups or distracter points on response time. The overall pattern of visual search was similar to the one found in

Experiment #3: response time increased with number of distracter groups and decreased with the number of distracter points. There was, however, no difference in that pattern between interactivity conditions. Visual search dynamics, therefore, could not have caused the difference in performance between the 1DOF and 3DOF interactivity conditions. Also, in a per-participant analysis of visual dynamics parameters, there was no clear clustering regarding the significance value of the effect that the number of distracters groups has on response time. On a general theoretical level, that casts suspicion on the dichotomous nature of the pop-out effect. However, it could be just a result of lack of statistical power in a single-participant single-DOF analysis.

A major difference between 1DOF and 3DOF was found in another area: the analysis of block order effects. During the experiment, every participant did two blocks of trials, one with 1DOF and another with 3DOF of interactive control of the viewpoint. Half of the participants did the 1DOF condition first, another half started with the 3DOF condition. Block order analysis has shown that the 3DOF block took significantly longer when it was the first block presented; 1DOF trials did not show a similar pattern of learning facilitation. When presented second, 3DOF block took approximately the same time as either of the 1DOF blocks. It is possible to conclude that 3DOF interactivity, rather than being more difficult inherently, is more difficult to master.

Here, as in previous experiments (#2 and #3), the crucial importance of learning for performance in interactive tasks is shown. In addition to block order effect, order of trials in each block was a uniformly significant factor: performance improved with practice. The same importance of practice, but on a larger scale, could be seen with the three outstandingly well-performing participants (CN, OS and PR), all of whom had extensive previous experience with tasks involving either rotating 3D information visualisations (PR,CN) or pattern recognition in scatterplot-like displays (OS).

This expertise was seen in the pen-and-paper tests as well: CN is exceptionally good in Mental Rotation Test and OS in Raven APM. No other link was found between the results of these tests and the performance in the main task. Thus the second question asked at the design stage of this experiment, whether performance is mediated by general cognitive and specific spatial abilities, was left unanswered.

Wand rotation data were recorded during each trial. These data, in conjunction with data on participants' performance and the results of pen-and-paper tests, could, in principle, provide answers to several important questions, such as whether performance is mediated by behaviour (i.e. wand movement patterns) or whether behaviour is mediated by the cognitive profile. Currently there are no widely accepted methods of 3D interaction analysis, although possible approaches are suggested here.

Several avenues of enquiry can be proposed for this research. One is to collate the data of vertical and horizontal rotations of the scene for each trial into a 2D frequency distribution; rotation around the depth axis can be discarded since it does not change the projection of the scene on the screen. This frequency distribution is a summary of the view angles that the stimuli were viewed under. At its most straightforward, it can then be presented as a 2D histogram, highlighting the angles of the scene that were used relatively more often. Figure 53 presents such a collation for each block, separately for each participant. It shows, for example, that in 1DOF condition, most, but not all, participants constrained their movements to 1DOF rotation. It also appears from that plot that 3DOF interactivity is constrained in a smaller range when it follows a block of 1DOF interactivity. To quantify that, a suitable metric of spread would have to be developed, possibly based on the 2D extensions of the concepts of second and fourth momentum (standard deviation and kurtosis).

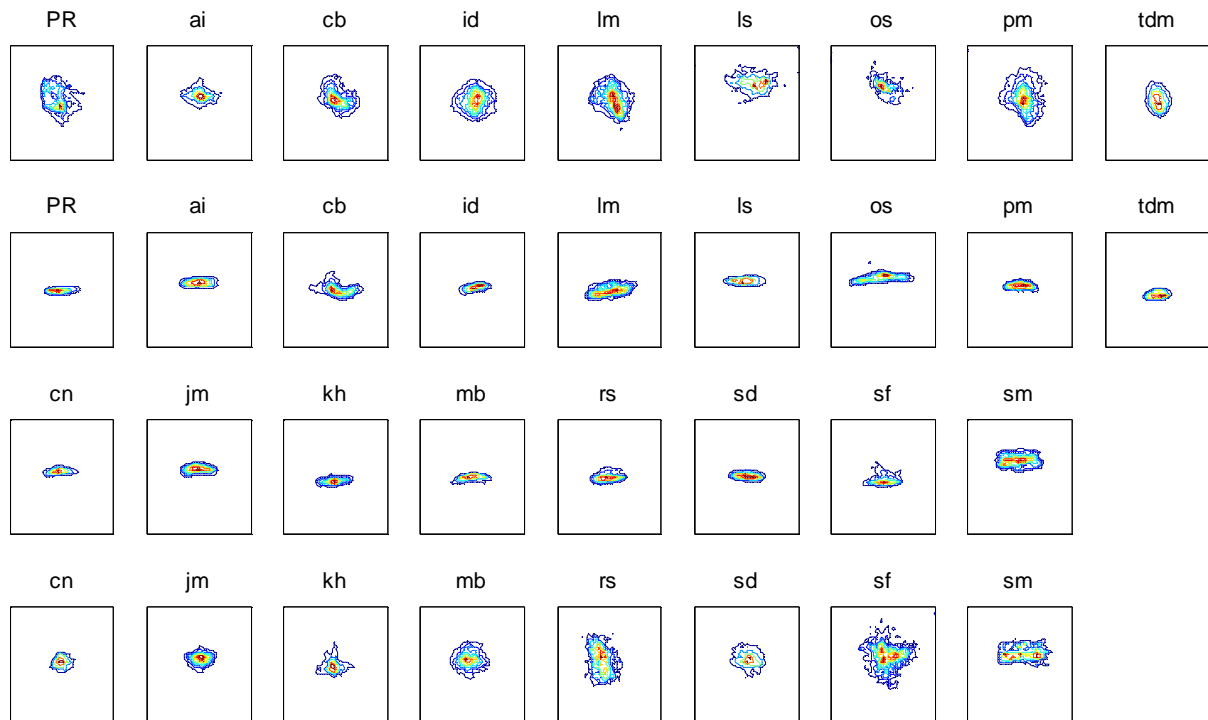


Figure 53 – 2D histograms of scene rotation per participant. First and fourth row: 3DOF; second and third: 1DOF; first and third row: block 1; second and fourth: block 2.

Another possible analysis of 3D interaction patterns is an extension of the analysis performed on the mouse logs in Experiments #2 and #3: dividing the rotation path into straight runs and sharp turns, and then comparing rotation patterns in terms of relative length and duration of straight runs or the sharpness of the turns. To do that for a 3D rotation path, time-interpolation will be necessary because of the way the data was recorded (a reading every time the angle changed for more than 0.01 radians). However, because of the gimbal lock problem²⁴, angular data can only be meaningfully interpolated in quaternion format. Thus, rotation data will have to be recoded into the quaternion format and then interpolated to create a constant-frequency reading. Only then may a derivation of rotation along the time dimension be done, providing a metric for the sharpness of turn.

²⁴ Gimbal lock is a condition that sometimes occurs when Euler angles (rotations around x, y, and z axes) are used to describe rotations; because of the topology of mapping from Euler angle space to a sphere, there are changes in rotation that can not be described as a change in Euler angles.

From the above brief outline, it is clear that the amount of additional research required for a comprehensive analysis of the wand rotation data is well beyond the scope of this thesis. It is, however, a fascinating topic and will be hopefully addressed in the future. For now, it suffices that the regression analysis and the paired comparison of response times have answered the main question that this experiment sets out to address.

Conclusions

Experiment #4 attempted to address the following questions: in which way degrees of freedom in interactive rotation affect performance; and what effect general reasoning and mental rotation skills have on performance. The findings provide a definitive answer for the first question only: restricted interactivity is easier to learn and thus preferable for non-expert users. There were no conclusive answers for the second and third questions.

Chapter 10. Discussion and conclusions

At the end of Chapter 1, the aim of this thesis was defined as follows: to produce novel findings in three related areas of Visual Analytics research: applied, methodological and theoretical. The theoretical aim was to lay the groundwork for a theory of visualisation comprehension. The methodological aim was to test the applicability of a cognitive psychology toolset to the evaluation of visualisation techniques. At the end of Chapter 4, the ‘pop-out’ methodology (Treisman, 1982) was chosen as the most appropriate tool. The applied aim was to assess the efficiency of a specific visualisation technique; in Chapter 5 the technique was decided to be an interactive 3D scatterplot.

To achieve these three aims, a series of successively planned experiments was undertaken, starting with a preliminary questionnaire study (Experiment #0, described in detail in Chapter 3). The results of Experiment #0 allowed for formulation of a framework for the objective psychophysical testing of visualisation efficiency, based on a juxtaposition of several theories of perception and comprehension. This framework was implemented and tested in a series of psychophysical experiments (Experiment #1 to #4, described in Chapters 6 to 9 respectively). Some of these experiments, in addition to testing the hypotheses that they were set out to investigate, brought to light topics that were not expected at the outset of this thesis, such as the importance of learning.

In the current chapter, all of the above will be discussed in detail. First, the empirical evidence will be summarised, with a special emphasis on the implications of the preliminary study results. Then, the success of this thesis in achieving each one of the three aims set out in Chapter 1 will be discussed, followed by sections on the additional topics that were highlighted by the empirical evidence, concluding with a proposal of options for future research.

Summary of Experimental results

The experimental evidence presented in previous chapters consisted of one preliminary study (Experiment #0) and four psychophysical experiments (#1-#4). Experiment #0 was a semi-structured interview designed to investigate visualisation practices in academic research. It identified the commonality in the purposes of exploratory data visualisation. From the responses, scientific data is most commonly visualised in order to detect outliers, and, to a lesser extent, clusters and trends. Based on these findings, a cognitive account of data visualisation comprehension emerged (Chapter 4), and a focused experimental programme was planned. The psychophysical experiments that followed studied the effects of interactivity on comprehension of 3D scatterplots. The experiments employed uniform methodology, measuring reaction time and relating it to the number of elements in stimulus array.

Experiment #1 sought to prove the overall feasibility of applying the ‘pop-out’ methodology to dynamic stimuli. The experiment produced conclusive results, showing that outlier detection in a constantly rotating 3D scatterplot does not exhibit the ‘pop-out’ effect, i.e. it takes more time to detect the outlier with more points in a scatterplot. Another important outcome of Experiment #1 was that it allowed for methodological improvements in experiment design: for example, the change from touch screen to a mouse as an input device; and the need for dissociation between conflicting interpretations of the cognitive processes involved (local movement comparison or 3D scene reconstruction). All the improvements were comparatively minor and were easily implemented in subsequent experiments, significantly improving the overall validity of this thesis.

Experiment #2 focused on the effects of interactivity in 3D scatterplot outlier detection, using improved stimuli and better data analysis methods. The results showed a significant benefit of interactive over automatic rotation. This benefit is, however, quantitative rather than qualitative, since the overall pattern of serial visual search for the target was present

regardless of interactivity. An unexpectedly strong learning effect was found during the sets of trials with interactive view control: interaction time decreased from trial to trial. In fact, the major benefit of interactivity seemed to be in that it gave the participants an opportunity to adapt the scene view and rotation speed to their own preference and thus improve their performance over the course of the experiment. In contrast, blocks of non-interactive trials did not show a comparable improvement in performance.

A limitation of Experiment #1 is that its results can be interpreted in two different ways. It is possible to hypothesise that the increase in number of elements chiefly affects motor planning of the response rather than perception of the outlier, and the results of Experiment #1 do not allow for dissociation between these two hypotheses. In Experiment #2, the analysis of mouse logs allowed separation of each trial into three stages: before view control (wait time); during view control (interaction time); and after view control and until the response (response time). This analysis showed that number of elements in the stimulus array affects solely interaction time, but not wait time or response time; it is therefore the perception of the outlier and not the motor control that is affected.

In Experiment #3 the task was changed from outlier detection to structure detection. The experiment addressed the theoretical question of objecthood and hierarchies of perception, in an attempt to decide between two contradictory accounts of visualisation comprehension, one derived from the Feature Integration Theory, and another from the Reverse Hierarchy Theory. The results showed that the elements that affect interaction time are high-level (point structures) rather than low-level (single points). This supports the Reverse Hierarchy Theory account of visualisation comprehension, and, by inference, validates the use of ‘pop-out’ methodology in visualisation comprehension evaluation.

Experiment #4 continued the line of investigation of previous experiments, addressing questions of 3D interactivity (as in Experiment #2) in the task of structure detection (as in Experiment #3). It was conducted in the virtual reality suite (HIVE) and made use of a fully

3D interaction device, the InterSense wand. The results indicated that full 3D interactivity significantly decreases performance. As in Experiment #2, the analysis of trial-to-trial changes in performance showed the importance of learning dynamics.

Applied recommendations

The applied focus of this thesis, defined in Chapter 5, was twofold: first, to test the efficiency of a 3D scatterplot in common Visual Analytics tasks; and second, to assess the efficiency of several different ways of presenting the depth information in the scatterplot using relative motion. Regarding the first question, it seems that it depends on the nature of the task: Experiments #1 and #2 seem to indicate that 3D scatterplot is not an efficient visualisation technique for outlier detection; Experiments #3 and #4 show that 3D scatterplot can be an efficient tool for structure detection, especially if the number of data points is large and the number of juxtaposed data groups is low. There is limited anecdotal evidence that 3D scatterplots with interactive view control can still be an efficient tool for outlier detection, in that they allow the user to quickly navigate to the view in which an outlier is clearly visible in 2D. However, since that possibility was actively disabled in the design of Experiments #1 and #2, any empirical support for this would have to come from further research.

The second applied question was the relative efficiency of different modes of interactive or non-interactive control of the viewpoint. Three modes were compared in the four experiments: non-interactive single-axis rotation; interactive single axis rotation; and full 3D interactive rotation. Of these, interactive rotation around a single vertical axis was the most efficient. Analysis of view rotation patterns in Experiment #3 shows that smaller and faster rotations tend to lead to better performance. This could be due to the similarity of the resulting visual sensation to sensations caused by minute head motions made when examining novel objects. Again, more evidence is necessary to test that hypothesis, e.g. an experiment that will study, using head and hand tracking, natural head movements during

exploration of complex physical objects; similar research has been done regarding head movements during locomotion (Grossman, Leigh, Abel, Lanska, & Thurston, 1988).

Previous research on the comprehension of 3D scatter plots (Wickens, et al., 1994) found that 3D presentation improves performance in questions with ‘high dimensional integration’ (i.e. requiring combination of data from several points and several axes to answer) but not with ‘low dimensional integration’ (e.g. questions regarding a position of a single point along one of the axes). Despite major differences in methodology, these findings are similar to the difference found in this thesis between outlier detection and trend detection. However, the same research does not find a benefit of user-induced scene rotation for surface perception in 3D, even when compared to planar presentation. One possible reason is that the camera control used in that research was unconstrained rotation around all axes, which, as found in Experiment #4, is less efficient than constrained rotation around a single vertical axis. Another reason could be that, since 1994, advances in graphic presentation capabilities, as well as in the 3D skills of the participants, have changed considerably. A more recent study, undertaken in a Virtual Reality environment, has demonstrated that when participants learn abstract 3D object structure by actively rotating the objects, the objects are recognized faster during a subsequent recognition task than when object structure was learned through passive observation (James, et al., 2002).

Overall, the recommendation for visualisation design that can be derived from this thesis is that 3D scatterplots with interactive single-axis rotation can be effective for trend detection in tri-variate data sets with hundreds of data points. Further research would extend to different visualisation types e.g. link-charts²⁵, as well as other modes of interactivity. Research on scene rotation by head position tracking (Arthur, et al., 1993) suggests that head tracking might be a promising technique for visualising 3D structures. Recent software

²⁵ link-chart is a generic visualisation technique representing relation data as glyphs connected by lines; link-charts are used to visualise mathematical graphs, network diagrams, mind-maps, UML diagrams, social networks and many more.

packages that track head position from camera video feed, such as faceAPI²⁶, make head tracking commercially feasible rather than an esoteric and expensive hardware setup.

Methodological considerations

Throughout this thesis, the ‘pop-out’ methodology was used to determine the cognitive processes involved in visualisation comprehension. This methodology consists of measuring the time it takes to identify an odd-one-out in a group of similar objects, while modulating the overall number of objects in this group. A constant or linearly increasing relationship between number of objects and the response time is considered valid within this methodology. All other interactions, e.g. decreasing or non-linear, would indicate an invalid result or the inapplicability of that methodology to a given task area. As outlined in Chapter 4, several theories of perception (Hochstein & Ahissar, 2002; Kubovy & Pomerantz, 1981; Pomerantz, 2006; Pomerantz, et al., 1977) tie the results of ‘pop-out’ experiments with the formation of high-level semantic parsing of the scene, or, in other words, to the origins of information. Within these theories, outlier identification time that is invariant with the number of objects is an indicator of implicit and automatic extraction of required information, whereas linearly increasing time indicates an explicit serial search for the target. While the ‘pop-out’ effect has been extremely well researched in relation to the basic principles of perception and information processing, the application of this effect to the applied context of assessing the efficiency of a visualisation technique is novel and unique.

The results of using ‘pop-out’ methodology in Experiments #1-#4 support its applicability to Visual Analytics tasks. That is despite the fact that the design of the experiments pushed the methodology far beyond the limits of its customary time-frame: in previous studies, the task never takes more than a second of passive perception (Kubovy & Pomerantz, 1981), whereas in this thesis the tasks consisted of 10-15 seconds of active interaction. The results

²⁶ <http://www.seeingmachines.com/product/faceapi/>

of all the experiments were valid within the paradigm. In all of them except one there was a clear linear relationship between the number of objects and the response time; the exception was the colour outlier trials of Experiment #2, where, consistent with previous research (Pomerantz, 2006), a ‘pop-out’ effect, i.e. constant response time not affected by the number of objects, was observed. Thus there were no empirical indications of non-validity or inapplicability of the methodology. Also, most of the numerous positive controls (situations for which a strong prediction exists) introduced into experimental design were confirmed. For example, in Experiment #1, outliers positioned closer to the distracters were more difficult to detect than the outliers positioned farther away; in Experiment #3 adding random small displacement to points in a target group made it more difficult to detect.

A unique feature of using ‘pop-out’ effect as the criterion for assessment of visualisation efficiency is its stand-alone quality: it allows evaluating a visualisation technique on its own, generating both qualitative and quantitative assessment of a visualisation. The qualitative assessment is the dichotomy of whether the technique is efficient, with the target ‘popping out’ at the user, or not, and the user has to search through the visualisation element by element. The quantitative assessment consists of the time constant for ‘pop-out’, and the linear parameters of constant and slope for visual search. These parameters can be compared across techniques, and are objective measures of performance.

In contrast, in numerous previous studies of visualisation efficiency where similar data (response time for correct answers) were collected (Arthur, et al., 1993; Healey, 1996; Healey & Enns, 2002; Healey, Tateosian, Enns, & Remple, 2004; Huber & Healey, 2005; Laidlaw, et al., 2005), these data are invariably used to compare several visualisation techniques to each other. Comparison studies, while valid and valuable in themselves, pose several scalability problems. Considering more than a few items becomes prohibitive in time, since extending the study to an additional technique requires comparing it to all the previous ones. Also, the sole result of a comparison study is an order relation in a set of

techniques, which could result in a non-linear ordering (A is better than B, B is better than C, but C is better than A) that is difficult to interpret.

Another assessment paradigm that is stand-alone is the comparison of human performance to a performance of an ‘ideal observer’ (Pelli & Farell, 1999). While producing valuable insights (e.g. Harris & Parker, 1994; Shovman & Ahissar, 2006), this approach is not applicable to interactive tasks, since building a model of an interacting ideal observer requires a better understanding of the visuo-motor system than is currently available. This prevents using the ‘ideal observer’ paradigm in real-life Visual Analytics scenarios.

In this thesis, the validity of the methodology is based on the Reverse Hierarchy Theory, tying the ‘pop-out’ effect to the formation of high-level mental representations of the scene, i.e. information. This theory, while promising, is relatively novel, and while ‘pop-out’ effect as a phenomenological observation is valid and replicable, associating it with specific cognitive mechanisms could be argued against. However, the methodology is relevant to the question of visualisation efficiency regardless of the Reverse Hierarchy Theory, as a minimal requirement criterion: a visualisation technique should *at least* allow effortless outlier detection, it being the most basic and common Visual Analytics task.

Between the ubiquity of outlier detection in Visual Analytics and a rigorous methodology of performance assessment based on the ‘pop-out’ effect, the methodology presented in this thesis offers a viable approach to visualisation quality evaluation. Further research, especially validation against other quality criteria, needs to be carried out, but already in this thesis, ‘pop-out’ methodology can be seen to yield useful results.

The science of visualisation

The main premise of the ‘theory of visualisation comprehension’ outlined in Chapters 2 and 4, is that the information extracted from visualisation and the high-level semantic parsing of the stimulus are essentially the same concept under different names. A corollary from the

Reverse Hierarchy Theory is that semantic parsing can be generated either by implicit ‘bottom-up’ processes or by explicit ‘top-down’ serial search, exhibiting distinct patterns of behaviour that can be observed using ‘pop-out’ methodology. An implication from the theories of comprehension (Kintsch, 1998) is that the information is formed by equally satisfying the constraints of stimulus and context, i.e. the task and by specific user’s abilities, skills, memories, emotional state etc.

This account is scientifically valid in a sense that it generates refutable hypotheses (Popper, 2002), for example, that performance in common information retrieval tasks would conform to patterns of perceptual organisation, or that changes in task or user would influence performance. In the experiments that comprised the empirical part of this thesis, predictions based on these hypotheses were formulated and tested. The first hypothesis, that the performance in common information retrieval tasks conforms to the patterns of perceptual organisation, is supported by the results of all of the experiments (see previous section). The effect of task could be observed in the difference in performance parameters depending on whether the task was to detect a colour outlier, position outlier or a trend. Lastly, high between-subject variability, and especially the superior performance of a small group of expert participants in Experiments #3 and #4, supported the claim that user background influences the ease with which information can be retrieved from a visualisation. Therefore it seems that, so far, the theory provides a feasible explanatory framework for visualisation research. However, further theoretical engagement would only improve the quality of the research in the field.

Effects of cognitive profile and of learning

Two phenomena were observed that were not in agreement with the theoretical framework. First, and contrary to predictions, there was no evidence for the effect of general cognitive skills on performance. Secondly, an unexpected strong positive effect of learning on performance was found. The effect of general cognitive skills on Visual Analytics tasks was

found by several studies in different paradigms (Domik & Gutkauf, 1994; Keehner, et al., 2008). For example, a study investigating learning the shape of complex 3D objects either by actively rotating them or by passively by observation, demonstrated that participants with low mental rotation score (Vandenberg & Kuse, 1978) benefit from interaction, whereas those with intermediate or high score do not (Meijer & van den Broek). The lack of evidence found in this thesis is most probably due to the specific choice of experimental design. The overall approach of this thesis focused on the uniformity of mental processes across users, in line with the general psychophysical approach (Donders, 1969) and based on the uniformity of visual insights found during the preliminary study. The design used in all four experiments was aimed mainly at disclosing across-subject effects from per-trial measurements; testing one participant yielded hundreds of measurements for the main analysis of ‘pop-out’ effect, and only one measurement of general cognitive skills.

A different series of experiments has to be run in order to adequately study the effects of user profile on visualisation comprehension. Such experiments could, for example, compare learning profile to cognitive profile; use a larger number of participants; and run a more comprehensive battery of tests (e.g. Wechsler, 1997). In the light of recent criticism that digit span is not a valid measure of general intelligence, alternative tests such as reading or operation span (reading sentences or solving equations aloud while trying to remember unrelated words) can also be considered (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002). Also, visualisation design based on modelling specific user abilities (Domik & Gutkauf, 1994) is a promising concept which could provide a complementary approach to the generalised one used here.

The other unexplained effect, that of learning, was seen both in the improvement of performance during the experiments, and in the exceptionally good performance of supposedly over-learned experts. The account put forward in Chapter 4 does not involve mechanisms for learning new tasks to the point where they exhibit ‘pop-out’ effect. There are, however, several models, based on a hierarchical approach to perception, which

accommodate for perceptual learning. A parsimonious account was proposed by Movshon and Kiorpes (2008), based on separate numerous findings showing that the maturation of visual function continues for a much longer time for complex visual functions and in higher visual cortical areas; that the locus of learning lies in these higher cortical areas; and that learning is most extensive for complex visual tasks. The authors hypothesise that visual maturation and perceptual learning reflect the same underlying processes: a cascade of developmental periods, ending early in life in lower visual areas, continuing longer in higher ones, and extending throughout life in yet higher ones.

Combining the Reverse Hierarchy Theory with the model of Movshon and Kiorpes (2008) creates the following model of learning for visualisation comprehension: at the higher, more abstract levels of perceptual hierarchy, the constraints of task and memory play a larger role in determining the semantic parsing and grouping of element and new concepts can be learned more easily. At the lower levels on the other hand, the critical window for learning (Levi, 2005) closes early in child development, making it impossible to change the gestalts formed at this stage by later learning, attention, intention or context. This hypothesis is in line with research in neural plasticity (e.g. Ahissar, 1999) but will, of course, have to be empirically tested in future studies.

Learning and interaction

Perceptual learning, as outlined in the previous section, can explain the improvement of performance in the experiment. However, the analysis of interaction and learning suggests a different explanation. In Experiment #2, participants' performance was shown to have been improved only in interactive trials, and in Experiment #4, manipulating the amount of interactivity available was shown to affect performance. This indicates a change in interaction strategy, either deliberate or inadvertent. Supporting evidence can be inferred from mouse movement patterns analysis: faster side to side sweeps correlated with higher speed and accuracy. The correlation could be explained by causal relation from either side.

On the one hand, faster sweeps could be a genuinely more efficient strategy. On the other hand, it is entirely possible that participants tend to hurry through easy trials, while slowing down and scrutinising the scene in harder ones. A support for the former explanation comes from the analysis of change in mouse movement over the course of the experiment. Mouse sweeps tend to become faster as the experiment progresses and performance increases, suggesting some degree of strategic learning.

More evidence is required to decide between these explanations. In any case, it is clear that interaction patterns such as sweep time affect both comprehension and learning in Visual Analytics tasks. Similar results were obtained in a series of experiments examining the effects of interactive visualizations and spatial abilities on a complex 3D task (Keehner, et al., 2008). In these experiments, the pattern of interactions made by best-performing participants was recorded and then shown, non-interactively, to another group of participants. The result was that non-interactive participants who watched optimal movements of the display performed as well as interactive participants who manipulated the visualization effectively and better than interactive participants who manipulated the visualization ineffectively. These findings, together with interaction logs from Experiments #2–#4, suggest a series of experiments, aiming to replicate the findings of Keehner et al. for 3D scatterplots. The results of such an experiment, i.e. a set of view control patterns optimal for perception, would have high applied relevance in the design of 3D Visual Analytics tools.

Pragmatics of Information Visualisation

In this last section of the discussion, the focus comes round to the first empirical results of this thesis, obtained during the preliminary interview study (Experiment #0). The recurring theme in all the interviews that were conducted was that there are only a few reasons to visualise data during the data analysis stage of the research. This presented a singularly parsimonious account spanning diverse visualisation techniques and research areas.

Apparently, a wide range of tools and techniques are used by researchers in different fields for a very limited number of aims, the most common of which is to identify outliers in the data. It seems that, just as data is transformed into a visualisation, a task in data domain is transformed, by the same transform function, into a visual task: finding a pattern in the data means finding a pattern in the chart. Experiment #0 results suggest that these visual tasks are limited in number and are uniform regardless of the subject area of the data and task.

The above finding was facilitated by the choice of methodology: using semi-structured interviews; only methodological differences precluded other studies to come across this fact earlier. Specifically, close-ended questionnaires (Robinson, 2009) do not allow the interviewees to expand on the reasons why the data were visualised; focusing on complex visual analysis systems tailor-made for specific task and audience (Coutaz, et al., 1995) prevents generalised findings about the fundamental building blocks of these systems.

Generally, that aspect of Visual Analytics research, the reasons for visualising data, is more often than not overlooked in the literature. Instead, most studies of visualisation focus either on the basic building blocks of a visualisation, such as geometries, layout, and colours, or on user interactivity and experience (Ware, 2000). Borrowing a terminology from a related field of linguistics, all these studies could be classified as focusing on the *syntax* and *morphology* of visualisations. In the same terms, the findings of Experiment #0 relate to the *pragmatics* of visualisation (Leech, 1983). The research into the pragmatics of speech is relatively recent, with the first studies appearing in the 1960s, but it has already revolutionised the approach to studying human language (Austin, 1962).

An account of visualisations in terms of pragmatics can benefit the science of visualisation, in particular Visual Analytics. A parsimonious set of pragmatic reasons for visualising data is an appropriate framework for research as well as for the design of new tools. However, a more in-depth study than the preliminary one conducted as part of this thesis is of course necessary, particularly since Experiment #0 has a very limited number of interviewees.

Several other pragmatics were suggested to the author in personal communication during presentations in national and international forums. One suggestion was that visualisations are useful for identifying recurring patterns in the data, and that this task is perceptually significantly different from identifying trends or clusters. That is consistent with the view held by some researchers (Julesz, 1981) that a texture is a gestalt in its own right. Another suggested pragmatic, made in relation to a very specific visualisation technique, the link-chart, was identifying connectivity, or gaps in connectivity, in a network. Mainly, though, the researchers taking part in the discussions agreed that the three reasons identified during Experiment #0 are indeed the most common aims of visualising data. In some works, the terminology is different, e.g. “*identifying commonalities and anomalies*” (Csinger, 1992) while still relating to the same pragmatic reasons: clusters, trends and outliers.

In general, there is a trend in research to move away from null hypothesis significance testing (Nickerson, 2000) to more elaborate exploration of experimental data. Multivariate data analysis (Tabachnik & Fidell, 2006) coupled with the capacity of real-time interactive visual exploration of the data (Bown, et al., 2010) offer an answer to a long-standing argument for a deeper engagement with the data (Tukey, 1977). Visually aided and statistically supported completion of the three tasks mentioned above essentially constitutes the full cycle of comprehensive data analysis in many disciplines: from preliminary data screening (outliers), to identifying separate groups of cases (clusters), and finally understanding the interrelationships and patterns within these groups (trends).

Conclusion

This thesis started from a need voiced by many in the Visual Analytics community: a need for valid and objective visualisation quality criteria. A gap in theoretical understanding was then identified, and a framework solution formulated. ‘Pop-out’ methodology was used as a criterion for information retrieval efficiency, and yielded several non-trivial results: that 3D scatterplots are efficient for structure detection, but not for outlier detection; that rotation of

the view around vertical axis is preferable to full 3D rotation; that interactivity improves performance by enabling learning; and more. Perhaps the main novelty is the formulation of an operative theory of information, and the three basic pragmatics of visualisation: outliers, clusters and trends. This offers a framework for further research: assessing immediacy of information retrieval in various visualisation techniques such as link-charts; looking for pragmatics in addition to the ones already found, e.g. patterns, links, and gaps, etc. Perhaps the most fascinating next question, after learning more about pragmatics, is how it interacts with aesthetics in the flow from data to visualisation to information; how the beauty of information either shines through the design of the visualisation, or is obscured by it.

Appendix: abstracts of peer-reviewed publications

Use of ‘pop-out’ paradigm to test graph comprehension in a three-dimensional scatter plot

This poster presentation, based on the results of Experiment #1, was made at the European Conference in Visual Perception 2008, Utrecht, and the abstract published in *Perception* (Shovman, Szymkowiak, Bown, & Scott-Brown, 2008). See the poster on the enclosed CD.

Abstract

The emerging field of Visual Analytics applies abstract data visualisation to analyse complex, multivariate data. Data visualisations comprise a full spectrum of pictorial and symbolic elements; thus, juxtaposing theories of visual perception and reading comprehension, graph comprehension can be linked to high-level perceptual organisation of the visual scene. The latter can be quantitatively assessed, e.g., by the ‘pop-out’ paradigm: constant response times with increasing stimulus array size. We assessed outlier detection in slowly rotating three-dimensional scatter plots, exploring comprehension of simple information – ‘an odd one out’. The outlier was defined by kinetic depth; stimulus arrays of several sizes were used. Results indicate that in larger stimulus arrays response times were longer while accuracy decreased, consistent with processes of visual search and not ‘pop-out’. In line with previous research, these results suggest that 3D charts, while visually impressive, are not efficient enough for data analysis and decision-making. Using ‘pop-out’ paradigm allows probing of high-level processes of graph comprehension with psychophysical methods, making it a viable approach to assessing data visualisation efficiency.

Inexpensive psychophysics of complex tasks: a case of mouse-log analysis

This poster presentation, based on the methodology of Experiment #2, was made at the Psychology Postgraduate Action Group conference, 2008, Manchester. See the poster on the enclosed CD.

Abstract

Experiments in visual psychophysics typically use highly specialised (and quite expensive) hardware and software, such as eye trackers, dedicated displays and video cards, response boxes, experiment-driving software etc. All these provide stimuli and data of high precision

and resolution. However we demonstrate that with a little effort, good results can be obtained with use of ubiquitous office-grade hardware and free software. In our experiment, participants view and interact with a rendering of a three-dimensional scatter plot. Participants' task is to identify an outlier in either colour or depth. In the latter case, it is necessary to interact with the scene in order to reconstruct the three-dimensional structure of the scatter-plot. The interactivity consists of rotating the scene either passively (starting/stopping a constant rotation) or actively (using a slider bar to control rotation angle). The experiment assesses the effect of different interaction modes on perception of depth in abstract three-dimensional scenes. Both the interaction and the response are made with a USB mouse; all mouse actions are logged and time-stamped. This allows us to obtain response timing with 1msec resolution and 20msec precision; post-experiment analysis of full mouse logs includes data cleanup, as well as separation of interaction (i.e. scene rotation) time from response (i.e. movement-to-target) time. Using standard office equipment also minimizes apparatus learning overhead. This rich analysis of mouse-logs demonstrates a simple low-cost method of measuring performance in complex interactive tasks, with acceptable time resolution and without using specialised equipment.

Changing the View: towards the theory of visualisation comprehension

This talk, based on the results of Experiment #2, was presented at 13th International Conference in Information Visualisation, 2009, Barcelona. A paper (see the enclosed CD) was published in the proceedings (Shovman, et al., 2009).

Abstract

The core problem of the evaluation of Information Visualisation is that the end product of visualisation – the comprehension of the information from the data – is difficult to measure objectively. This paper outlines a description of visualisation comprehension based on two existing theories of perception: Principles of Perceptual Organisation and the Reverse Hierarchy Theory. The resulting account of the processes involved in visualisation comprehension enables evaluation that is not only objective, but also non-comparative, providing an absolute efficiency classification. Finally, as a sample application of this approach, an experiment studying the benefits of interactivity in 3D scatterplots is presented.

Is visual search a high-level phenomenon? Evidence from structure perception in 3D scatterplots

This talk, based on the results of Experiment #3, was presented the Applied Vision Association Easter Meeting, 2010, Liverpool.

Abstract

Increasing use of 3D scatterplots for trend detection in Visual Analytics raises the theoretical question: what constitutes a single perceptual object in such a task? Previously we have shown that detection of a 3D position of a single outlier point exhibits characteristics of serial visual search. From this, Feature Integration Theory would conclude that every point is a complex perceptual object, and that detection of trends or patterns in a scatterplot would be a serial search based on the number of constituent points. Conversely, the Reverse Hierarchy Theory predicts that the object in a trend detection task would be a highest-level task-relevant arrangement, i.e. a group of points. Therefore, it would predict that increasing the number of point groups would increase the difficulty of the task, and make no prediction of the effect of increasing the number of points. Participants' task was to identify, in a 2-4AFC task, a group of points whose positioning exhibited a 3D structure when they were rotated. Number of points per group (64, 100, 144, 196 or 256) and the number of groups (2, 3 or 4) were manipulated independently. The dependent variable was the scene rotation duration, i.e. the time when the 3D structure was visible. We found that rotation time increased with number of point groups and decreased with number of points per group. This is consistent with Reverse Hierarchy Theory and contradicts Feature Integration Theory. In conjunction with previous experiments, these results support connecting processes of visual search to task-relevant high-level semantics of the scene, rather than to its low-level visual features.

References

- Ahissar, M. (1999). Perceptual learning. *Current Directions in Psychological Science*, 8(4), 124. doi: 10.1111/1467-8721.00029
- Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: specificities to stimulus characteristics. *Vision Research*, 36(21), 3487-3500. doi: 10.1016/0042-6989(96)00036-3
- Amende, N. (2010). *A Structured Review of Information Visualization Success Measurement*. Paper presented at the 24th BCS Conference on Human Computer Interaction, University of Abertay Dundee. <http://hci2010.abertay.ac.uk/accepted.html>
- Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, 61(3), 303.
- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, 27(1), 17-21. doi: 10.2307/2682899
- Arkfeld, M. R. (2001). *The digital practice of law: a practical reference applying technology concepts to the practice of law*: Law Partner Publishing.
- Arthur, K. W., Booth, K. S., & Ware, C. (1993). Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems (TOIS)*, 11(3), 239-265. doi: 10.1145/159161.155359
- Atran, S. (1993). *Cognitive foundations of natural history: Towards an anthropology of science*: Cambridge University Press.
- Attfield, S., Blandford, A., Dowell, J., & Cairns, P. (2008). Uncertainty-tolerant design: Evaluating task performance and drag-and-link information gathering for a news-writing task. *International Journal of Human-Computer Studies*, 66(6), 410-424. doi: 10.1016/j.ijhcs.2007.12.001
- Austin, J. L. (1962). *How to do things with words*. Boston, MA: Harvard University Press.
- Bertin, J. (1967). *Sémiologie graphique*. Paris: Mouton
- Bigwood, S., & Spore, M. (2003). *Presenting Numbers, Tables, and Charts*. Oxford: Oxford University Press.
- Black, E., Floridi, L., & Third, A. (2009). Introduction to the special issue on the nature and scope of information. *Synthese*, 1-3.
- Blake, R., & Sekuler, R. (2005). *Perception* (5 ed.). New York: McGraw-Hill.
- Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. *ACM SIGGRAPH Computer Graphics*, 11(2), 192-198. doi: 10.1145/965141.563893
- Bown, J., Fee, K., Sampson, A., Shovman, M. M., Falconer, R., Goltsov, A., et al. (2010). *Information visualization and the arts-science-social science interface*. Paper presented at the IITM'10, Allahabad, India.
- Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: an introduction to design, data analysis, and model building*. New York: Wiley.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*: The MIT Press.
- Brown, L. (1993). *The New Shorter Oxford English Dictionary* (Vol. 1 & 2). Oxford: Oxford University Press.
- Casselmann, B. (2000). Pictures and proofs. *Notices of American Mathematical Society*, 47(10), 1257-1266.
- Chang, D., & Nesbitt, K. V. (2006). Developing Gestalt-based design guidelines for multi-sensory displays. *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop*, 57, 9-16
- Chang, D., Nesbitt, K. V., & Wilkins, K. (2007a). *The Gestalt Principle of Continuation Applies to both the Haptic and Visual Grouping of Elements*. Paper presented at the EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.

- Chang, D., Nesbitt, K. V., & Wilkins, K. (2007b). *The gestalt principles of similarity and proximity apply to both the haptic and visual grouping of elements*. Paper presented at the Australasian conference on User Interface.
- Chen, C. (2004). *Information Visualization: Beyond the Horizon*. London; New York: Springer.
- Chen, C., Härdle, W., & Unwin, A. (2008). *Handbook of data visualization*: Springer Verlag.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531-554. doi: 10.2307/2288400
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30(2), 163-183. doi: 10.1016/S0160-2896(01)00096-4
- Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., & Young, R. (1995). *Four easy pieces for assessing the usability of multimodal interaction: the CARE properties*. Paper presented at the INTERACT.
- Cramer, J. S. (2003). The origins and development of the logit model. Retrieved from http://www.cambridge.org/resources/0521815886/1208_default.pdf
- Csinger, A. (1992). The psychology of visualization. Retrieved from <ftp://ftp.cs.ubc.ca/local/techreports/1992/TR-92-28.pdf>
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. *Perception of Space and Motion*, 5, 69-117.
- Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-computer interaction*. Harlow, England: Prentice Hall.
- Domik, G. O., & Gutkauf, B. (1994). *User modeling for adaptive visualization systems*. Paper presented at the IEEE Conference on Visualization.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412. doi: 10.1016/0001-6918(69)90065-1
- Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, 80(1-2), 61-95. doi: 10.1016/S0010-0277(00)00151-7
- Du Toit, S. H. C., Steyn, A. G. W., & Stumpf, R. H. (1986). *Graphical exploratory data analysis*. New York: Springer-Verlag.
- Fekete, J. D., & Plaisant, C. (2002). *Interactive information visualization of a million items*. Paper presented at the 2002 IEEE Symposium on Information Visualization (InfoVis 2002).
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-392.
- Floridi, L. (2005). Semantic conceptions of information. *Stanford Encyclopedia of Philosophy*.
- Freedman, E., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. *Diagrammatic representation and inference*, 59-141.
- Friendly, M., & Denis, D. J. (2001). Milestones in the history of thematic cartography, statistical graphics, and data visualization. Retrieved from <http://www.datavis.ca/milestones/>
- Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological science*, 7(4), 237. doi: 10.1111/j.1467-9280.1996.tb00366.x
- Gibson, J. J. (1986). *The ecological approach to visual perception*: Lawrence Erlbaum.
- Goldstein, E. B. (2009). *Sensation and perception* (8 ed.): Wadsworth Cengage Learning.
- Gombrich, E. H., & Gombrich, E. H. (1995). *The story of art*: Phaidon.

- Graham, M., Kennedy, J., & Benyon, D. (2000). Towards a methodology for developing visualizations. *International Journal of Human-Computer Studies*, 53(5), 789-807. doi: 10.1006/ijhc.2000.0415
- Green, T. M., Ribarsky, W., & Fisher, B. (2009). Building and applying a human cognition model for visual analytics. *Information Visualization*, 8(1), 1-13. doi: 10.1057/ivs.2008.28
- Grinstein, G. G., & Smith, S. (1990). *Perceptualization of scientific data*. Paper presented at the SPIE.
- Grossman, G. E., Leigh, R. J., Abel, L. A., Lanska, D. J., & Thurston, S. E. (1988). Frequency and velocity of rotational head perturbations during locomotion. *Experimental Brain Research*, 70(3), 470-476. doi: 10.1007/bf00247595
- Harris, J. M., & Parker, A. J. (1994). Constraints on human stereo dot matching. *Vision Research*, 34(20), 2761-2772. doi: 10.1016/0042-6989(94)90231-3
- Healey, C. G. (1996). *Choosing effective colours for data visualization*. Paper presented at the IEEE conference on Visualization, San Francisco, CA.
- Healey, C. G., & Enns, J. T. (2002). Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Transactions on Visualization and Computer Graphics*, 5(2), 145-167. doi: 10.1109/2945.773807
- Healey, C. G., Tateosian, L., Enns, J. T., & Remple, M. (2004). Perceptually based brush strokes for nonphotorealistic visualization. *ACM Transactions on Graphics (TOG)*, 23(1), 64-96. doi: 10.1145/966131.966135
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, 45(13), 1707-1724. doi: 10.1016/j.visres.2004.12.021
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791-804. doi: 10.1016/s0896-6273(02)01091-7
- Huber, D. E., & Healey, C. G. (2005). *Visualizing data with motion*. Paper presented at the 16th IEEE conference on Visualization, Minneapolis, Minnesota.
- Interrante, V. (2000). Harnessing natural textures for multivariate visualization. *IEEE Computer Graphics and Applications*, 20(6), 6-11. doi: 10.1109/mcg.2000.888001
- James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, 34(3), 383. doi: 10.3758/BF03195466
- Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802), 91-97. doi: 10.1038/290091a0
- Kanizsa, G. (1976). Subjective contours. *Scientific American*, 234(4), 48. doi: 10.1038/scientificamerican0476-48
- Kant, I., & Meiklejohn, J. M. D. (1882). *Critique of pure reason*: G. Bell.
- Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science*, 32(7), 1099-1132. doi: 10.1080/03640210801898177
- Keil, A., Muller, M. M., Ray, W. J., Gruber, T., & Elbert, T. (1999). Human gamma band activity and perception of a gestalt. *Journal of Neuroscience*, 19(16), 7152-7161.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*: Cambridge Univ Pr.
- Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception and Psychophysics*, 52(1), 18-36. doi: 10.3758/BF03206757
- Kosslyn, S. M. (2006). *Graph Design for the Eye and Mind*. Oxford: Oxford University Press.
- Kubovy, M., & Pomerantz, J. R. (1981). *Perceptual Organization*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Kuhn, T. S. (1996). *The structure of scientific revolutions*: University of Chicago Press.

- Laidlaw, D. H., Kirby, R. M., Jackson, C. D., Davidson, J. S., Miller, T. S., Da Silva, M., et al. (2005). Comparing 2D vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics*, 59-70. doi: 10.1109/TVCG.2005.4
- Langdridge, D., & Hagger-Johnson, G. (2009). *Introduction to research methods and data analysis* (2nd ed.): Pearson Education (US).
- LaViola Jr, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, 32(1), 47-56. doi: 10.1145/333329.333344
- LeDoux, J. (1996). *The emotional brain*. New York.
- Leech, G. N. (1983). *Principles of pragmatics*: Longman Pub Group.
- Levi, D. M. (2005). Perceptual learning in adults with amblyopia: A reevaluation of critical periods in human vision. *Developmental Psychobiology*, 46(3), 222-232. doi: 10.1002/dev.20050
- Li, L., Peng, T., & Kennedy, J. (2010). *A rule based taxonomy of dirty data*. Paper presented at the International Academic Conference on Data Analysis, Data Quality and Metadata Management, Singapore.
- Lidal, E. M., Langeland, T., & Giertsen, C. (2007). A Decade of Increased Oil Recovery in Virtual Reality. *Computer Graphics and Applications, IEEE*, 27(6), 94-97. doi: 10.1109/MCG.2007.141
- Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. *Bioscience*, 51(5), 341-352. doi: 10.1641/0006-3568
- Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2), 110-141. doi: 10.1145/22949.22950
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*: Henry Holt and Co., Inc. New York, NY, USA.
- McCandless, D. (2009). *Information is Beautiful: The Information Atlas*: Collins.
- Meijer, F., & van den Broek, E. L. (2010). Representing 3D virtual objects: Interaction between visuo-spatial ability and type of exploration. *Vision Research*, 50(6), 630-635. doi: 10.1016/j.visres.2010.01.016
- Movshon, J. A., & Kiorpes, L. (2008). Perceptual learning and the critical period for visual development. *Perception*, 37, 166.
- Necker, L. A. (1832). Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. *The London and Edinburgh Philosophical Magazine and Journal of Science*, 1(5), 329-337.
- Nesbitt, K. V. (2004). *MS-Taxonomy: A conceptual framework for designing multi-sensory displays*. Paper presented at the Eighth International Conference on Information Visualisation.
- Nickerson, R. S. (2000). Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy* 1. *Psychological Methods*, 5(2), 241-301. doi: 10.1037/1082-989X.5.2.241
- Norman, D. A., & Bernbach, H. A. (1970). *Models of human memory*: Academic Press.
- Norman, D. A., & Shallice, T. (2000). Attention to action: Willed and automatic control of behavior. *Cognitive neuroscience: a reader*, 376-390.
- Nowell, L., Hetzler, E., & Tanasse, T. (2001). *Change blindness in information visualization: A case study*. Paper presented at the IEEE Symposium on Information Visualization, San Diego, CA.
- Pashler, H., & Yantis, S. (2004). *Stevens' Handbook of Experimental Psychology, Sensation and Perception*: Wiley.
- Peck, R., Olsen, C., & Devore, J. L. (2008). *Introduction to statistics and data analysis*: Duxbury Pr.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of Optical Society of America A*, 16(3), 647-653. doi: 10.1364/JOSAA.16.000647
- Perls, F., Hefferline, R. F., & Goodman, P. (1951). *Gestalt therapy*. New York (auch: New York, 1965).

- Pinker, S. (1990). A theory of graph comprehension. *Artificial intelligence and the future of testing*, 73-126.
- Playfair, W. (1801). *The Commercial and Political Atlas*. London.
- Polys, N. F., Kim, S., & Bowman, D. A. (2007). Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. *Computer Animation and Virtual Worlds*, 18(1), 19-38. doi: 10.1002/cav.159
- Pomerantz, J. R. (2006). Colour as a Gestalt: Pop out with basic features and with conjunctions. *Visual Cognition*, 14(4), 619-628. doi: 10.1080/13506280500195052
- Pomerantz, J. R., Sager, L. C., & Stoeber, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3(3), 422-435. doi: 10.1037/0096-1523.3.3.422
- Popper, K. R. (2002). *Conjectures and refutations: The growth of scientific knowledge*: Psychology Press.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86(3), 446-461. doi: 10.1037/0033-2909.86.3.446
- Raven, J. C., Raven, J., & Court, J. H. (1962). *Advanced progressive matrices: Sets I and II*: HK Lewis London.
- Remez, R. E., Fellowes, J. M., Pisoni, D. B., Goh, W. D., & Rubin, P. E. (1998). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication*, 26(1-2), 65-73. doi: 10.1016/S0167-6393(98)00050-8
- Robinson, A. C. (2009). *Needs Assessment for the Design of Informatics Synthesis Visual Analytics Tool*. Paper presented at the IEEE conference on Information Visualisation, Barcelona.
- Rubin, E. (1921). *Visuell wahrgenommene Figuren: Studien in psychologischer Analyse*: Gyldendalske boghandel.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80(1-2), 1-46. doi: 10.1016/S0010-0277(00)00152-9
- Scott-Brown, K. C. (Cartographer). (1996). The perception of motion [PhD Thesis].
- Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: implications for instruction. *Educational Psychology Review*, 14(1), 47-69. doi: 10.1023/A:1013180410169
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication* (Urbana, IL: University of Illinois Press).
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701.
- Shovman, M. M., & Ahissar, M. (2006). Isolating the impact of visual perception on dyslexics' reading ability. *Vision Research*, 46(20), 3514-3525. doi: 10.1016/j.visres.2006.05.011
- Shovman, M. M., Szymkowiak, A., Bown, J. L., & Scott-Brown, K. C. (2008). Use of 'pop-out' paradigm to test graph comprehension in a three-dimensional scatter plot. *Perception*, 37 (ECVP Abstract Supplement), 79.
- Shovman, M. M., Szymkowiak, A., Bown, J. L., & Scott-Brown, K. C. (2009). *Changing the View: towards the theory of visualisation comprehension*. Paper presented at the IEEE conference on Information Visualisation, Barcelona.
- Sollenberger, R. L., & Milgram, P. (1993). Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(3), 483-499.
- Steinman, R. M., Pizlo, Z., & Pizlo, F. J. (2000). Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution. *Vision Research*, 40(17), 2257-2264. doi: 10.1016/S0042-6989(00)00086-9
- Tabachnik, B., & Fidell, L. S. (2006). *Using Multivariate Statistics*. New York: Allyn & Bacon.

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society Press.

Tory, M., & Moller, T. (2004). *Rethinking visualization: a high-level taxonomy*. Paper presented at the IEEE Symposium on Information Visualization, Austin, Texas, USA.
<http://doi.ieeecomputersociety.org/10.1109/INFVIS.2004.59>

Tovee, M. J. (1998). Is face processing special? *Neuron*, 21(6), 1239-1242. doi: 10.1016/S0896-6273(00)80644-3

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2), 194-214. doi: 10.1037/0096-1523.8.2.194

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. doi: 10.1016/0010-0285(80)90005-5

Troscianko, T., Gregory, L., Hawker, K., Bhatt, A., Porter, G., Lovell, G., et al. (2007). *The big, the bad, and the ugly: an investigation into the effect of video screen size on visual cognition*. Paper presented at the The Twelfth AVA Christmas Meeting, Aston University, Birmingham.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Van Wijk, J. J. (2005). *The value of visualization*. Paper presented at the 16th IEEE conference on Visualization, Minneapolis, MN.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599.

Ware, C. (2000). *Information Visualization: Perception for Design*. San Francisco; London: Morgan Kaufman Publishers, Inc.

Ware, C., Arthur, K., & Booth, K. S. (1993). *Fish tank virtual reality*. Paper presented at the INTERACT'93 and CHI'93 conference on Human factors in computing systems.

Ware, C., & Knight, W. (1995). Using visual texture for information display. *ACM Transactions on Graphics (TOG)*, 14(1), 3-20. doi: 10.1145/200972.200974

Ware, C., & Osborne, S. (1990). *Exploration and virtual camera control in virtual three dimensional environments*. Paper presented at the Symposium on Interactive 3D graphics

Watson, J. D. (2001). *The double helix: A personal account of the discovery of the structure of DNA*: Touchstone.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale-III (WAIS-III). *San Antonio, TX: The Psychological Corporation*.

Wertheimer, M. (1938). Laws of organization in perceptual forms. *A source book of Gestalt psychology*, 71-88.

Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). Implications of graphics enhancements for the visualization of scientific data: Dimensional integrality, stereopsis, motion, and mesh. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(1), 44-61.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception and Psychophysics*, 63(3), 381-389. doi: 10.3758/BF03194406

Wyszecki, G., & Stiles, W. S. (1967). *Color science*: Wiley New York.

Zaitsev, A. V., & Skorik, Y. A. (2002). Mathematical description of sensorimotor reaction time distribution. *Human Physiology*, 28(4), 494-497. doi: 10.1023/A:1016550503294

Zakia, R. D. (1975). *Perception and photography*: Prentice Hall.

Zhu, Y. (2007). Measuring effective data visualization. *Advances in Visual Computing*, 652-661.